

Expectation values, experimental predictions, events and entropy in quantum gravitationally decohered quantum mechanics

Bernard S. Kay and Varqa Abyaneh

Department of Mathematics, University of York, York YO10 5DD, UK

E-mail: (address for correspondence) bsk2@york.ac.uk

Abstract. We restate Kay’s 1998 hypothesis which simultaneously offers an objective definition for the entropy of a closed system, a microscopic foundation for the Second Law, a resolution of the Information Loss (and other) Black-Hole Puzzle(s) as well as an objective mechanism for decoherence. The hypothesis presupposes a conventional unitary theory of low-energy quantum gravity and offers all this by taking the physical density operator of a closed system to be the partial trace of its total density operator (assumed pure) over gravity and by defining its physical entropy to be its ‘matter-gravity entanglement entropy’. We further recall Kay’s 1998 modification of non-relativistic (many-body) quantum mechanics based on Kay’s hypothesis with a Newtonian approximation to quantum gravity. In this modification, we find formal expectation values for certain ‘observables’ such as momentum-squared and Parity are altered but those for functions of positions are unaltered. However, by arguing that every real measurement can ultimately be taken to be a position measurement, we prove that, in practice, it is impossible to detect any alteration at all and, in particular, we predict no alteration for an experiment recently proposed by Roger Penrose. Nevertheless, Kay’s modification contains no Schrödinger Cat-like states, and also allows a possible ‘events’ interpretation which we tentatively propose and begin to explore. We also obtain a first result of Second-Law type for a non-relativistic toy-model closed system and argue that similar results will apply for a wide class of model Newtonian and post-Newtonian closed systems although we argue that ordinary actual laboratory-sized systems can never be treated as closed for the purpose of calculating their entropy. Compared with ‘collapse models’ such as GRW, Kay’s Newtonian theory does a similar job while being free from ad hoc assumptions.

PACS numbers: 03.65.Yz, 03.65.Ta, 04.60.-m, 04.70.Dy

1. Introduction

1.1. Background

In 1998, [1], one of us (BSK) argued that several of the problems and puzzles concerning decoherence and thermodynamic behaviour, which are inherent in our current understanding of quantum physics in general, as well as several problems and

puzzles specifically related to quantum black holes <i>‡, would appear to find natural resolutions if one makes the following hypothesis (which is in three parts): First, that for the resolution of these problems and puzzles, a quantum gravitational setting is required; more precisely, a low-energy quantum gravitational setting, where, ‘low’ here refers to energies well below the Planck energy. Second, that low energy quantum gravity is a quantum theory of a conventional type with a given closed quantum gravitational system described by a total Hilbert space $\mathcal{H}_{\text{total}}$ which arises as a tensor product

$$\mathcal{H}_{\text{total}} = \mathcal{H}_{\text{matter}} \otimes \mathcal{H}_{\text{gravity}} \quad (1)$$

of a matter and a gravity Hilbert space <ii>, and with a total time-evolution which is unitary, but that, while, as would usually be assumed in a standard quantum theory, one still assumes that there is an ever-pure time-evolving ‘underlying’ state, modelled by a density operator of form

$$\rho_{\text{total}} = |\Psi\rangle\langle\Psi|, \quad (2)$$

$\Psi \in \mathcal{H}_{\text{total}}$, at each ‘instant of time’, one should add the new assumption that the *physically relevant* density operator is not this underlying density operator, but rather its partial trace, ρ_{matter} , over $\mathcal{H}_{\text{gravity}}$. Third, that the physical entropy of a closed quantum gravitational system is to be identified with the von Neumann entropy of ρ_{matter} ,

$$S_{\text{physical}} = -k \text{tr}(\rho_{\text{matter}} \ln \rho_{\text{matter}}).$$

In other words, <iv>:

The physical entropy of a closed system is its matter-gravity entanglement entropy.

With this hypothesis, an initial underlying state of a closed quantum gravitational system with a low degree of matter-gravity entanglement would be expected to become more and more entangled as time increases and thus the physical entropy, as we have defined it above, would be expected to increase monotonically, thus (when the theory is applied to a model for the universe as a whole) offering the possibility of an objective microscopic explanation for the Second Law of Thermodynamics and (when the theory is applied to a model closed system consisting of a black hole sitting in an otherwise empty universe) offering a resolution to the Information Loss Puzzle. It offers both of these things in that, by defining the physical entropy of ρ_{total} to be the von-Neumann entropy of ρ_{matter} , one reconciles an underlying unitary time-evolution on $\mathcal{H}_{\text{total}}$ with an entropy which varies/increases in time <v>.

A second paper, [2] (also by BSK) also in 1998, investigated the implications of the hypothesis of [1] for the decoherence of ordinary matter in the ‘Newtonian’ – i.e. non-relativistic, weak-gravitational-field – regime. With some further assumptions, the conclusions of this work were that, if, in ordinary non-relativistic quantum mechanics, the centre-of-mass degree of freedom of a uniform-mass-density ball of mass M and

‡ The small Roman numerals in angle-brackets refer to the section entitled ‘Notes’ (Section 10) at the end of this article.

radius R would be described by a Schrödinger wave function $\psi \in L^2(\mathbb{R}^3)$ then the physically relevant density operator ρ_{matter} (from now on, where it can cause no confusion, we will often just call this ρ) is given by

$$\rho(\mathbf{x}, \mathbf{x}') = \rho_0(\mathbf{x}, \mathbf{x}') e^{-D(\mathbf{x}, \mathbf{x}')}, \quad (3)$$

where $\rho_0(\mathbf{x}, \mathbf{x}') = \psi(\mathbf{x})\psi^*(\mathbf{x}')$ is the position-space density matrix one would expect to have were gravitationally-induced-decoherence effects to be ignored, and D , which is called the *decoherence exponent*, is a certain function of \mathbf{x} and \mathbf{x}' which depends on \mathbf{x} and \mathbf{x}' only through the square of their difference $(\mathbf{x} - \mathbf{x}')^2$ and which vanishes when $\mathbf{x} = \mathbf{x}'$. Explicitly, D is given (below and from now on, a stands for $|\mathbf{x} - \mathbf{x}'|$ and we regard D as a function of a) by

$$D(a) = 216M^2 \int_0^\infty \frac{(\sin(\kappa) - \kappa \cos(\kappa))^2}{\kappa^7} \left(\frac{\kappa a/R - \sin(\kappa a/R)}{\kappa a/R} \right) d\kappa,$$

and we remark that this is a monotonically increasing function of a [2].

To summarize, the overall effect of the new theory is to replace the ‘would-be’ Schrödinger wave function ψ (or more precisely the would-be density operator $\rho_0 = |\psi\rangle\langle\psi|$) by the physical density operator ρ of (3).

Moreover, in [2], an asymptotic limit of D was obtained for the case where $a \ll R$,

$$D(a) \simeq \alpha a^2 + O(\alpha^2 a^4 \ln(a/R)), \quad (4)$$

where (by default we shall assume Planck units where $G = c = \hbar = k = 1$) $\alpha = 9M^2/R^2$. We refer to (4) as the *Gaussian* asymptotic regime.

We remark that we expect this Gaussian regime to always give a good approximation for a ball with a mass around or bigger than the Planck mass, since then, whenever a fails to be very much less than R , $e^{-D(a)}$ will anyway be very small.

We also remark that, as long as the ball in question is far from being on the verge of collapsing to form a black hole (which we anyway expect it to need to be for our Newtonian approximation to be valid) then the constant α in the Gaussian regime (4) will (in Planck units) be very much less than one.

An asymptotic limit was also obtained in the case where $a \gg R$,

$$D(a) \simeq 24M^2 \ln\left(\frac{a}{R}\right) + O(1). \quad (5)$$

in view of which we refer to this limit as the *logarithmic* asymptotic regime.

We shall sometimes study a model one-dimensional analogue of (3) which is taken to be exactly Gaussian, so that the relation between the physical and would-be density operator is given by

$$\rho(x, x') = \rho_0(x, x') e^{-\kappa(x-x')^2}, \quad (6)$$

where

$$\rho_0(x, x') = \psi(x)\psi^*(x').$$

where ψ is taken to be a ‘would-be’ wavefunction on the line and κ is a positive constant. We shall call this the *one-dimensional Gaussian model*.

In view of (4) and the subsequent remark, we would expect that, if we identify κ in (6) with $\alpha = 9M^2/R^2$, then, for M around or bigger than the Planck mass, this one-dimensional Gaussian model will give a good approximate description of the relation between the would-be wavefunction and the physical density operator describing the centre-of-mass degree of freedom of a uniform mass bead (i.e. a uniform mass ball with a narrow straight tube bored through its middle) of mass M and radius R constrained to slide on a straight wire, where x is taken to denote the position of the centre-of-mass of the bead along the wire relative to some choice of origin.

We note that if such a bead is constrained to move between hard stops, say a distance $2(R + \delta)$ apart – so that, calling the position of the centre of the bead when it is half-way between the stops ‘the origin’, any would-be wave function, $\psi(x)$ for the centre-of-mass motion is required to vanish at $x = \pm\delta$ – then the Gaussian model would, again in view of (4), be expected to give a good approximation for any mass M provided $\delta \ll R$.

[2] also generalized the formula (3) to the many-body wave function (for the centre of masses) of a system of many balls. The general formula, for a system of N balls takes the form

$$\rho(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N) = \rho_0(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N) e^{-D(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N)}, \quad (7)$$

where $\rho_0(x_1, \dots, x_N; x'_1, \dots, x'_N)$ takes the form $\Psi(x_1, \dots, x_N) \Psi^*(x'_1, \dots, x'_N)$ for some many-real-variable wave function Ψ and D is a certain function which depends on $\mathbf{x}_1 \dots \mathbf{x}_N$ and $\mathbf{x}'_1 \dots \mathbf{x}'_N$ only through the distances $|\mathbf{x}_I - \mathbf{x}_J|$, $|\mathbf{x}'_I - \mathbf{x}'_J|$, $|\mathbf{x}_I - \mathbf{x}'_J|$ and $|\mathbf{x}'_I - \mathbf{x}_J|$ ($I, J = 1 \dots N$) and which vanishes whenever $\mathbf{x}_I = \mathbf{x}'_I$ for all $I = 1 \dots N$.

In much of this paper, we shall focus our attention on states involving a single ball in 3 dimensional space or on our even simpler one-dimensional Gaussian model but we note that in the generalization to the case of N balls of equal masses and radii, there is a generalization of the Gaussian asymptotic regime in the leading term of which D takes the explicit form <vii>

$$D(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N) = \alpha (\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_N - \mathbf{x}'_1 - \mathbf{x}'_2 - \dots - \mathbf{x}'_N)^2.$$

This leads us to define an N -body generalization of the one-dimensional Gaussian model by

$$\rho(x_1, \dots, x_N; x'_1, \dots, x'_N) = \rho_0(x_1, \dots, x_N; x'_1, \dots, x'_N) e^{-\kappa(x_1 + x_2 + \dots + x_N - x'_1 - x'_2 - \dots - x'_N)^2}$$

where again $\rho_0(x_1, \dots, x_N; x'_1, \dots, x'_N)$ takes the form $\Psi(x_1, \dots, x_N) \Psi^*(x'_1, \dots, x'_N)$ for some many-real-variable would-be wave function Ψ . We shall refer to this latter for one of the results obtained in the Appendix (Section 9). (Note <vii> also includes the generalization to the many-ball case of the logarithmic asymptotic regime.)

[2] further proposed a Schrödinger picture time-evolution rule appropriate to this Newtonian approximation, according to which (in the single-ball case) the time-evolving physically relevant density operator is obtained by adopting the formula (3) at each moment in time, t , so that

$$\rho(t; \mathbf{x}, \mathbf{x}') = \rho_0(t; \mathbf{x}, \mathbf{x}') e^{-D(\mathbf{x}, \mathbf{x}')}, \quad (8)$$

where $\rho_0(t; \mathbf{x}, \mathbf{x}') = \psi_t(\mathbf{x})\psi_t^*(\mathbf{x}')$ where the would-be wave function ψ_t evolves exactly as in ordinary non-relativistic quantum mechanics, for some choice of Hamiltonian, H , so that $\psi_t(\mathbf{x}) = \langle \mathbf{x} | e^{-iHt} | \psi_0 \rangle$ <viii>. (Similarly, in the many-ball case and the Gaussian model etc., we adopt formula (7), respectively (6) etc. at each moment in time, again assuming that the relevant would-be wave functions evolve in time according to some choice of Hamiltonian which we again call H .)

For a Schrödinger Hamiltonian of form $H = -\nabla^2/2M + V$, it may easily be checked that this amounts to the statement that $\rho(t; \mathbf{x}, \mathbf{x}')$ evolves according to the master equation

$$\dot{\rho} = \frac{i}{2M} (\nabla^2 - \nabla'^2 + V(\mathbf{x}) - V(\mathbf{x}')) \rho - \frac{1}{M} \nabla'_a D(x-x') (-i\nabla_a - i\nabla'_a) \rho. \quad (9)$$

We remark that it is easy to see that the analogous master equation for the one-dimensional Gaussian model can be written in the simple operator form

$$\dot{\rho} = -i[H, \rho] - \frac{2\kappa}{M} [x, [p, \rho]]. \quad (10)$$

While these master equations can be shown *not* to have ‘Gorini-Kossakowski-Sudarshan/Lindblad’ (cf. [6], [7]) form (we shall abbreviate this elsewhere to ‘GKS/L form’) and even not to arise from a time-indexed family (i.e. not necessarily semigroup) of ‘completely positive maps’ acting on an initial ρ (see <ix> for an explanation of all this) they clearly do have solutions, specified by the formulae (8), (6), which consist of a density operator <x> ρ_t at each instant of time and (in either case) the map from ρ_t to ρ_{t+s} arises in the form $\rho_{t+s} = \mu(s)\rho_t$ where the operators (on the space of density operators on our matter Hilbert space) $\mu(s)$ clearly form a one-parameter group (not just semi-group).

1.2. Purpose of present paper and summary of results

The project which gave rise to the present paper had three purposes: Firstly to explore the prospects for experimentally testing the theory of [1], secondly to explore whether/how the theory may form a basis for a solution to some of the conceptual problems of quantum mechanics (i.e. of the ‘measurement problem’), and, thirdly and lastly, to explore in more detail what the theory predicts about the time-dependence of entropy.

In the present paper, as a first step, we shall begin to explore all three of these issues, proceeding as if, for the relevant systems, the Newtonian approximation of [2] applies. It was argued in [2] that this assumption is unjustified for physically realistic laboratory-sized systems, at least as far as questions of entropy are concerned – the problem being that, while such systems will satisfy the necessary approximations in the sense that the relevant speeds are slow enough and the relevant mass distributions very far away from black-hole collapse etc., it is *not* expected (for the reasons listed in [2] and discussed further in Section 7) to be appropriate to treat them as *closed*. (They should rather be treated as open systems as indicated in Note <xii>.)

In this article, the systems we treat will nevertheless be treated as if they *were* closed. Thus considerable caution should be exercised in assigning a direct physical interpretation to our results. However, we shall argue in the later parts of our Discussion section (Section 7) that, as far as our first two questions (experiments and measurement problem) are concerned, the answers we obtain will, when properly interpreted as we explain in Section 4, survive when we drop this assumption and that, as far as our third question (entropy) is concerned, the results we obtain will at least survive as a part – albeit perhaps a small part – of the full results for actual (presumably partly relativistic) closed systems and also be of interest as providing ‘toy models’ for these full results.

Proceeding then as if the relevant systems were closed and our Newtonian theory of [2] were exact, we amplify on each of these issues in turn: [2] pointed out that, as far as the mathematical formulae are concerned, the Newtonian theory of [2] leads to large effects in the ‘amount of decoherence’ already for ‘small macroscopic’ systems. Thus e.g. a ball with 10 times the Planck mass with a would-be wavefunction for its centre-of-mass a Schrödinger Cat-like superposition

$$\psi = c_1\psi_1 + c_2\psi_2 \quad (11)$$

(cf. (67) in Note <vi>) where ψ_1 and ψ_2 are sharply localized around two distinct locations separated by a distance a equal to one tenth of the ball’s radius, will, by (4), have a decoherence exponent, $D(a)$, of approximately 9, and thus have a physical density operator very close to that for a statistical mixture

$$\rho = |c_1|^2|\psi_1\rangle\langle\psi_1| + |c_2|^2|\psi_2\rangle\langle\psi_2| \quad (12)$$

of states localized at the two locations. More precisely ρ will be as in (68) with $\langle g_1|g_2\rangle$ approximately equal to e^{-9} . It thus may seem at first sight plausible that, with suitable experimental setups, differences from the predictions of standard quantum mechanics should be experimentally detectable in the laboratory. Moreover, there have been several proposals (the proposals we mention below are [8], [9]) for such experimental setups, albeit the theories which the authors of these setups had in mind were somewhat different from the theory of [1], [2]. In order to explore what the prospects are for such experimental tests, we need of course to take a view about how our mathematical formalism relates to experiment. To this end, we have explored the consequences of adopting what we shall call here the *naive pragmatic interpretation* according to which the usual sorts of ‘observables’, e.g. (for the single ball model) position (\mathbf{x}) and momentum (\mathbf{p}) and functions of these, parity (\mathcal{P}) and so on, are represented by the usual self-adjoint operators and the expectation value of the observable represented by the self-adjoint operator A obtained in the state ρ is given by the usual formula

$$\text{The expectation value, } \langle A \rangle_\rho, \text{ of } A \text{ in the state } \rho \text{ is equal to } \text{tr}(\rho A). \quad (13)$$

As we report in Section 2 we find that, with this naive pragmatic interpretation, the expectation values of \mathbf{x} and \mathbf{p} and also of arbitrary functions of \mathbf{x} are unchanged,

$$\text{tr}(\rho\mathbf{x}) = \text{tr}(\rho_0\mathbf{x}), \quad \text{tr}(\rho\mathbf{p}) = \text{tr}(\rho_0\mathbf{p}), \quad \text{tr}(\rho f(\mathbf{x})) = \text{tr}(\rho_0 f(\mathbf{x})), \quad (14)$$

but that expectation values for \mathbf{p}^2 and parity, \mathcal{P} , acquire extra terms (see (16) and (18)). In Section 3, we explore the change in the naive expectation values for \mathbf{p}^2 further and point out that these lead to a non-zero state-independent minimum value for the expectation value of \mathbf{p}^2 and, related to this, to a modification of the usual \mathbf{x} – \mathbf{p} Heisenberg uncertainty relation. (Section 3 stands to one side of the main line of the paper. The results are expected to be of interest in their own right but are not needed or referred to in subsequent sections.)

We shall however (i.e. in spite of the alterations in naive expectation values of some observables from the those of standard quantum mechanics) argue that our naive pragmatic interpretation is too naive. As we shall discuss in detail in Section 4, if one considers an experimental setup e.g. to determine the deviation of the expectation value of \mathbf{p}^2 for a ball in a particular physical state from its expectation value according to standard quantum mechanics in the corresponding would-be wave function, one would presumably need to introduce at least a second ‘probe’ particle and a measurement of \mathbf{p}^2 for the original ball might then e.g. amount to some sort of position measurement involving the probe particle. However, as long as our probe particle is also non-relativistic $\langle \mathbf{x} \rangle$, the two-body versions of our formulae (14) are assumed to apply and one can easily see from these that they entail, when formula (13) is applied at the two-body level, that

Any sort of position measurement on the relevant physical density operator will always give the same result as that predicted, according to standard quantum mechanics, for the same measurement on the corresponding would-be wave function.

In the case discussed above, the relevant physical density operator is the two-body density operator replacing the time-evolute of the tensor product of the would-be wave function for the ball and the initial would-be wave function for the probe particle. We will prove a generalization of this result for N -bodies in Section 4 and we shall refer to this latter generalization as our *Position Measurement Theorem*.

One might of course think of some cleverer setup in which the measurement on the probe particle is also not a sort of position measurement, but we shall advocate the view, that for a measurement to have been convincingly made, then, sufficiently far along the Heisenberg-von Neumann chain of measurements (see Chapter 6 in [10]) it eventually ends up, or can be regarded as ending up, as a position measurement – traditionally one talks of the position of a ‘pointer’ on a ‘dial’. We shall call this view the *corrected pragmatic interpretation*; it is still ‘pragmatic’ insofar as it still applies the formula (13) at *some* stage along the Heisenberg-von Neumann chain. Adopting this interpretation, the general N -body argument we mentioned above applies and we thus have the rather surprising conclusion:

On our corrected pragmatic interpretation, the experimental predictions of the Newtonian approximate theory of [2] are identical with those of standard quantum mechanics!

(A)

Of course we expect that, in the fully relativistic strong-field regime, the hypothesis of [1] will predict (large) differences but we expect them to depend on the post-Newtonian aspects of the theory. In Section 4 (see especially Note <xxvi>) we give an estimate which suggests that these post-Newtonian corrections will, however, be completely negligible for the sort of laboratory experiments which are currently under consideration.

Of particular interest in connection with our conclusion (A) above is the photon-interferometer-based experiment [21], [22] proposed by Roger Penrose to detect modifications to the predictions of standard quantum mechanics predicted by theories of ‘collapse-model’ type such as that of Ghirardi-Rimini-Weber [13] (elsewhere in the paper, we refer to this as the ‘GRW’ model) and others and, in particular, by the gravity-induced quantum state reduction proposal [20] [21] [22] of Penrose himself. At the heart of the experiment is a macroscopic object (crystal) attached to a movable interferometer mirror which, by the arrangement of the experiment would get put, according to a standard quantum-mechanics analysis, into a superposition, one branch of which remains at rest, while the other oscillates between two distinct locations. If (e.g. because of Penrose’s gravity-induced quantum state reduction) the crystal decoheres as a result of visiting (i.e. in one of the branches of the superposition) the distinct location during its oscillation cycle, a superposition of the photon state in the two arms of the interferometer will (by the design of the experiment – see Section 4.1) get converted into a partly decohered mixture thereby changing the predicted exit-paths of a certain fraction of the photons travelling through the interferometer from what would be predicted by standard quantum mechanics, and this change can then be detected with suitable photon detectors.

According to Kay’s theory [2] such a crystal will (if the size is sufficient and the distinct location suitably distant) in some ways similarly to what is predicted by Penrose’s quantum state reduction, be predicted to decohere each time it visits its distinct location and yet, according to our conclusion (A) (and assuming that it applies when some of the particles involved are photons <xi>) what is detected at the detector must be identical to what would be predicted by standard quantum mechanics.

In Subsection 4.1 we explain how this difference in predictions between the theory of Kay and the quantum state reduction of Penrose may be traced to the following qualitative difference between their proposed decoherence mechanisms. On the Penrose proposal it is (at least implicitly) assumed that once the crystal has decohered, then it cannot later recohere. In the Kay theory of [2], on the other hand, the superposition involving the crystal is understood to decohere when, in one of the branches of its superposition, the crystal is in its distinct location, but to recohere each time the crystal returns to being in the same location in both branches.

Turning to the second of our purposes, which, we recall, was to explore how the theory of [1] may form a basis for a solution to some of the conceptual problems of quantum mechanics, we begin by remarking that while the *corrected pragmatic interpretation* described above seems the best we can do at the moment to interpret

our theory and make experimental predictions, it is still unsatisfactory in that, like the standard interpretation of standard quantum mechanics, it makes essential reference to ‘observables’ and hence implicitly to ‘observers’ and, as e.g. John Bell eloquently argued (see e.g. [12]) one might hope that our current understanding of quantum mechanics will one day be superseded by a theory which is more objective in nature. Indeed many authors have asked for a modification of quantum mechanics (or maybe for a changed interpretation of standard quantum mechanics) in which there is an objective notion of ‘events’ which ‘happen’. For theories which are couched in terms of a time-evolving density operator $\rho(t)$, as in either the full hypothesis of [1] or the Newtonian approximate theory of [2] being considered here (but also ‘collapse models’ which we mentioned above in connection with the Penrose experiment)) a natural proposal would seem to be <xiv> <xv> <xvi>:

The set of possible events which can happen at a given time, t , is to be identified with the set of spectral subspaces of $\rho(t)$ at that time. The probability with which a given such event occurs is to be identified with $m\lambda$ where λ is the eigenvalue of $\rho(t)$ belonging to that subspace and m is its multiplicity (i.e. the dimension of the subspace).

(B)

As we will discuss further in Section 5, this proposed interpretation would, for example, interpret the state (12) in terms of two possible events, namely the subspace spanned by ψ_1 and the subspace spanned by ψ_2 , and assign to these events the probabilities $|c_1|^2$, $|c_2|^2$ respectively. We shall also explore, in Section 5, what this proposed interpretation entails for the Gaussian model (6) and shall actually simplify this further by studying that only to first order in κ .

In order to do this, we obtain, in the Appendix, the diagonalization of the density operator of the one-dimensional Gaussian model to first order in κ . A typical result is that, for a given would-be wave function ψ which is an even function or an odd function, the density operator ρ in (6) takes, to first order in κ , the form

$$\rho = \lambda_1 |\phi_1\rangle\langle\phi_1| + \lambda_2 |\phi_2\rangle\langle\phi_2|$$

where $\lambda_1 = 1 - 2\kappa\langle\psi|x^2\psi\rangle$ is (for small κ) close to 1 and ϕ_1 (see (33)) resembles ψ while $\lambda_2 = 2\kappa\langle\psi|x^2\psi\rangle$ is small and ϕ_2 (equal to a normalization constant times $x\psi$) has the opposite parity to ψ . According to our proposal, this would be interpreted to mean that, for such an (even or odd) would-be symmetric quantum wavefunction on the real line, there are two events, one with probability λ_1 which is identified with the subspace spanned by ϕ_1 and one with probability λ_2 which is identified with the subspace spanned by ϕ_2 .

Concerning our third purpose, which is to explore what the theory of [1] has to say about the time-evolution of entropy, we make a small beginning in Section 6 by exploring how the von-Neumann entropy S and also the entropy-like quantity $S_1 = 1 - \text{tr} \rho^2$ vary in time, mainly treating our one-dimensional Gaussian model restricted further to would-be wave functions which are either even or odd and to the regime where

working to first order in κ suffices so that we may utilize the diagonalization results obtained in the Appendix. As is apparent from the formulae quoted above for the eigenvalues λ_1 and λ_2 , the values of S and S_1 for this model when the system is in a given would-be state, ψ , at a given time t , will be correlated with the value of the standard-quantum-mechanical quantity $\langle \psi(t) | x^2 \psi(t) \rangle$ at that time, which, in standard quantum mechanics, is of course a measure of the ‘spreading of the (would-be) wave-packet’. This leads, if, as is the case for a free Hamiltonian, the Hamiltonian governing the would-be dynamics leads to spreading of the wave packet at times on either side of a time of minimum spread, to a ‘two-sided entropy increase’ result. On the other hand, for a would-be Hamiltonian, say, of harmonic-oscillator type, the would-be wavefunction will be periodic in time, and therefore the amount of wavepacket-spreading and therefore also the values of S and S_1 will be periodic in time, indicating a continual process of decoherence and recoherence. We point out that this is analogous to the decoherence-followed-by-recoherence mentioned above in our discussion of the Penrose experiment. We also compare and contrast this Hamiltonian-dependent behaviour of S and S_1 with the typical sort of monotonic entropy-increase result that one finds for density operators which satisfy master equations of GKS/L form which are usually the relevant master equations in collapse models. Specifically, we compare and contrast the results mentioned above which are relevant to the master equation (10) with an easily-derived result on the Hamiltonian-independent monotonic increase of entropy for the Barchielli-Lanz-Prosperi [27] master equation (38)

$$\dot{\rho} = -i[H, \rho] - c[x, [x, \rho]]$$

(c a positive constant) which may be regarded as a prototype master equation of GKS/L form. (Elsewhere in this paper, we shall refer to this as the ‘BLP’ equation.)

The main part of the paper ends, in section 7, with a discussion of various issues arising from the earlier sections.

We shall summarize here some of the main overall conclusions which we draw, in Section 7 by comparing and contrasting what we have learned about the theory of [2] with the properties of ‘collapse models’ such as GRW [13]. First, we conclude that, for ordinary laboratory-sized quantum systems, the cautionary remarks mentioned at the start of this subsection can be ignored as far as quantum mechanical measurements are concerned and (in view of our Position Measurement Theorem and our ‘corrected pragmatic interpretation’) our theory really does have identical experimental predictions to standard quantum mechanics (although there will be deviations when relativistic effects are taken into account) while, at the same time, being free from macroscopic Schrödinger Cat-like superpositions. This is to be compared and contrasted with collapse models which predict small deviations from the predictions of standard (non-relativistic) quantum mechanics while also being free from Schrödinger Cat-like superpositions. Another feature of collapse models is that they give rise to an increasing amount of decoherence as time increases. For our theory, we give, in Section 7, plausibility arguments for a two-sided entropy-increase result, of the sort discussed in Section

6 and mentioned above, for a wide variety of ‘generic’ model closed systems within the Newtonian framework of [2] and we also argue for a plausible post-Newtonian extension of such results. Concerning the two-sidedness, we advocate, in Section 7, the natural physical interpretation that the negative-times should be discarded as physically irrelevant and the time-zero density operator regarded as the ‘initial state’. (If the closed system is taken as modelling the universe, then this will model the initial physically relevant density operator of the universe.) In this sense, such results would amount to proofs, for appropriate models, of the Second Law. As far as decoherence is concerned, such Second-Law results can be interpreted as entailing an increasing amount of decoherence as time increases. However, we do not interpret these results as applying to actual ordinary (non-relativistic) systems of laboratory size because indeed we will argue in Section 7 that, as far as questions of entropy are concerned, actual closed systems must be partly relativistic and indeed may well be of galaxy-size! Rather, we argue that actual small ordinary laboratory-sized systems need to be regarded as *open* systems (See Note <xii>) and that the correct mechanism behind their decoherence is essentially the same as that advocated on the traditional environment-induced decoherence paradigm. (The entropy, i.e. amount of decoherence, will then increase or decrease with time according to the model and state etc.) (We have omitted here a comparison of approaches to ‘events’ in the two theories. For this, we refer to Sections 5 and 7 and especially Note <xv>.)

Our overall conclusion from this comparison is that our approximate Newtonian theory of [2] does a similar job to the job which ‘collapse models’ such as GRW [13] do – albeit in a slightly different and partly unexpected and surprising way but with the advantage that, while collapse models are ad hoc, our Newtonian approximate theory is part and parcel of a general hypothesis (i.e. the hypothesis of [1]) which also resolves a number of other puzzles.

This paper arose out of an ongoing research project of one of the authors (BSK). While both authors were involved with the new results presented in the paper, the main involvement of VA was with Sections 2, 3, and 9 with Section 3 and Subsections 9.1 and 9.2 being mainly due to VA. Some further material related to these sections can be found in [4]. Some parts of the rest of the paper (including the last paragraph of Subsection 1.2, the proposed interpretation in terms of events in Sections 1 and 5 and the *Position Measurement Theorem* discussed in Sections 1 and 4, including the discussion of the Penrose experiment in Subsection 4.1)) are mainly due to BSK, as are Sections 6 and 7.

The paper concludes with an extensive Notes section (Section 10, due to BSK) which was needed to put the results of the main body of the paper in context and clarify their significance. It both collects together some relevant background material and also includes some new material which is expected to be of interest in its own right.

In particular, aside from their importance as relevant background to the new results reported in this paper, the first paragraph of Subsection 1.1 together with Notes <i>, <iii>, <iv>, <v> and <xii> will, we hope, be of interest as constituting a self-contained, updated (i.e. in the light of more recent work of BSK) re-statement of

the hypothesis, which was first made in [1], for a resolution of the puzzles listed in Note <i>.</i>

2. Naive pragmatic interpretation: trace formulae

It is easy to see from (3) that, if A is any function, $f(\mathbf{x})$, of position and also if A is the momentum operator \mathbf{p} , then $\text{tr}(\rho A)$ is equal to $\text{tr}(\rho_0 A)$. To see the former, notice that

$$\text{tr}(\rho f(\mathbf{x})) = \int f(\mathbf{x}) \rho(\mathbf{x}, \mathbf{x}) d^3 \mathbf{x}$$

and that, by virtue of $D(\mathbf{x}, \mathbf{x}) = 0$,

$$\rho(\mathbf{x}, \mathbf{x}) = \rho_0(\mathbf{x}, \mathbf{x}).$$

For the latter, we combine

$$\text{tr}(\rho p_a) = \int -i \frac{\partial}{\partial x^a} \rho(\mathbf{x}, \mathbf{x}')|_{\mathbf{x}=\mathbf{x}'} d^3 \mathbf{x}'$$

and

$$-i \frac{\partial}{\partial x^a} \rho(\mathbf{x}, \mathbf{x}') = -i \left(\frac{\partial}{\partial x^a} \rho_0(\mathbf{x}, \mathbf{x}') \right) e^{-D(\mathbf{x}, \mathbf{x}')} - 2i\alpha \rho_0(\mathbf{x}, \mathbf{x}') (x^a - x'^a) e^{-D(\mathbf{x}, \mathbf{x}')} \\ + \text{higher order terms which vanish when } \mathbf{x} = \mathbf{x}'$$

(here we used (4)) and notice that all but the first term in the second equation will vanish when $\mathbf{x} = \mathbf{x}'$.

On the other hand, one easily sees ('a' here refers to a single index and is not summed over) from

$$-\frac{\partial^2}{\partial x^{a2}} \rho(\mathbf{x}, \mathbf{x}') = - \left(\frac{\partial^2}{\partial x^{a2}} \rho_0(\mathbf{x}, \mathbf{x}') \right) e^{-D(\mathbf{x}, \mathbf{x}')} - 4\alpha \left(\frac{\partial}{\partial x^a} \rho_0(\mathbf{x}, \mathbf{x}') \right) (x^a - x'^a) e^{-D(\mathbf{x}, \mathbf{x}')} \\ + 2\alpha \rho_0 e^{-D(\mathbf{x}, \mathbf{x}')} + \text{higher order terms which vanish when } \mathbf{x} = \mathbf{x}'$$

that, for each single Cartesian component of momentum squared,

$$\text{tr}(\rho p_a^2) = \text{tr}(\rho_0 p_a^2) + 2\alpha, \quad (15)$$

whereupon, summing over a , we clearly have

$$\text{tr}(\rho \mathbf{p}^2) = \text{tr}(\rho_0 \mathbf{p}^2) + 6\alpha. \quad (16)$$

It is interesting to note that all the above results remain unchanged if one replaces the term $e^{-D(\mathbf{x}, \mathbf{x}')}$ in (3) by its Gaussian asymptotic form $e^{-\alpha(\mathbf{x} - \mathbf{x}')^2}$ or even if one replaces the latter by its 'first-order in α ' form $1 - \alpha(\mathbf{x} - \mathbf{x}')^2$, albeit the resulting expression for ρ is then not to be taken seriously except when \mathbf{x} and \mathbf{x}' are sufficiently nearby for $\alpha(\mathbf{x} - \mathbf{x}')^2$ to be small compared to 1. In this spirit, we write

$$\rho_{\text{approx}}(\mathbf{x}, \mathbf{x}') = \rho_0(\mathbf{x}, \mathbf{x}') (1 - \alpha(\mathbf{x} - \mathbf{x}')^2)$$

i.e.

$$\rho_{\text{approx}} = \rho_0 - \alpha x^a x^a \rho_0 + 2\alpha x^a \rho_0 x^a - \alpha \rho_0 x^a x^a = \rho_0 - \alpha[x^a, [x^a, \rho_0]]$$

(summed) whereupon we have

$$\text{tr}(\rho_{\text{approx}} A) = \text{tr}(\rho_0 A) - \alpha \text{tr}([x^a, [x^a, \rho_0]] A) = \text{tr}(\rho_0 A) - \alpha \text{tr}(\rho_0 [x^a, [x^a, A]]) \quad (17)$$

(summed) where we have used the easy standard trace identity, $\text{tr}(A[B, C]) = \text{tr}([A, B]C)$, etc. in the last equality. One easily checks that this latter equation reproduces the exact expectation values for $f(\mathbf{x})$, \mathbf{p} and p_a^2 (unsummed) (and \mathbf{p}^2) found above.

On the ‘naive pragmatic interpretation’ discussed in the introduction, the equation $\text{tr}(\rho f(\mathbf{x})) = \text{tr}(\rho_0 f(\mathbf{x}))$ predicts e.g. an identical diffraction pattern to that of standard quantum mechanics e.g. for a double-slit experiment involving a ball of ordinary matter even as heavy as or heavier than the Planck mass (cf. the experiments of Zeilinger et al [19] albeit for smaller masses). On the other hand, if we consider the wavefunction for, say, the ground state of such a ball in a spherical or cubical box etc., then the results (15), (16) would seem to predict a detectable difference in the expectation value of the squared momentum.

One can also calculate $\text{tr}(\rho \mathcal{P})$ where \mathcal{P} is the parity operator.

$$\text{tr}(\rho \mathcal{P}) = \int \rho_0(-\mathbf{x}, \mathbf{x}) e^{-D(-\mathbf{x}, \mathbf{x})} d^3 \mathbf{x}$$

Assuming the above ‘first-order in α ’ approximation to the Gaussian approximation held (this would be the case e.g. for the centre-of-mass wavefunction of a ball with mass much bigger than the Planck mass and radius R confined to a cubical or spherical box of diameter not much bigger than $2R$) we would have, to first order in α

$$\text{tr}(\rho \mathcal{P}) = \text{tr}(\rho_0 \mathcal{P}) - \alpha \text{tr}(\rho_0 [x^a, [x^a, \mathcal{P}]]) = \text{tr}(\rho_0 \mathcal{P}) - 4\alpha \langle \psi | \mathbf{x}^2 \mathcal{P} \psi \rangle. \quad (18)$$

We remark that, if the would-be wave function $\psi(\mathbf{x})$ has even parity, then this is $1 - 4\alpha \langle \psi | \mathbf{x}^2 \psi \rangle$ i.e. 1 minus 4α times the squared uncertainty in x of the would-be wave function and, if $\psi(\mathbf{x})$ has odd parity, then it is $-1 + 4\alpha \langle \psi | \mathbf{x}^2 \psi \rangle$ i.e. -1 plus 4α times the squared uncertainty in x of the would-be wave function. Also, the counterpart to this result for our one-dimensional Gaussian model (6), is that, for a would-be wave function $\psi(x)$ on the line

$$\text{tr}(\rho \mathcal{P}) = \text{tr}(\rho_0 \mathcal{P}) - \kappa \text{tr}(\rho_0 [x, [x, \mathcal{P}]]) = \text{tr}(\rho_0 \mathcal{P}) - 4\kappa \langle \psi | x^2 \mathcal{P} \psi \rangle. \quad (19)$$

This would be appropriate e.g. to the bead-on-a-wire constrained to move between sufficiently close-together hard stops as discussed in Subsection 1.1. Moreover, if the wave function is even, we’ll have

$$\text{tr}(\rho P) = 1 - 4\alpha \langle \psi | x^2 \psi \rangle \quad (20)$$

(and if it is odd,

$$\text{tr}(\rho P) = -1 + 4\alpha \langle \psi | x^2 \psi \rangle). \quad (21)$$

We will discuss the significance of (18), (19), (20) and (21) further in Section 5.

3. Minimum $\Delta \mathbf{p}^2$ and modified uncertainty relations

(*Note* The results of this section are not referred to or needed in the remainder of the paper. Note also that, throughout this section, we shall not assume that a repeated index is summed over.)

Continuing to adopt the naive pragmatic interpretation, it is interesting to note that Eq. (15) implies that there is a minimum value (i.e. $2\alpha = 18M^2/R^2$) for the expectation value $\langle p_a^2 \rangle = \text{tr}(\rho p_a^2)$ of the square of each component of the momentum in a given physical state for the centre of mass motion of one of our balls of mass M and radius R . Given that $\text{tr}(\rho p_a) = \text{tr}(\rho_0 p_a)$, and defining the squared uncertainty, $(\Delta p_a)^2$, in p_a in the usual way by

$$(\Delta p_a)^2 = \text{tr}(\rho p_a^2) - \text{tr}(\rho p_a)^2,$$

we have

$$(\Delta p_a)^2 = \text{tr}(\rho_0 p_a^2) - \text{tr}(\rho_0 p_a)^2 + 2\alpha = (\Delta_0 p_a)^2 + 2\alpha \quad (22)$$

where $\Delta_0 p_a$ is the usual uncertainty in p_a of the corresponding would-be wave function. So the squared uncertainty in each component of momentum also has a minimum possible value of 2α .

To get some idea of the orders of magnitude involved, we can convert the minimum expectation value, 6α , of \mathbf{p}^2 into a ‘maximum wavelength’ (cf. the de Broglie relation ‘ $\lambda = 2\pi\hbar/|\mathbf{p}|$ ’), so that

$$\lambda = 2\pi \langle \mathbf{p}^2 \rangle^{-1/2} = \frac{2\pi}{\sqrt{6\alpha}} = \frac{2\pi R}{3\sqrt{6}M}.$$

In any system of units, this is $(2\pi/3\sqrt{6})(M_{\text{Planck}}/M)R$. Thus, for the centre-of-mass motion of a uniform density ball of mass around the Planck mass (i.e. around 10^{-5} g) λ will be of the same order of magnitude as the ball’s radius. For a proton, modelled as a uniform density ball with the proton mass and a radius around 10^{-13} cm, λ will be of the order of 10 kilometres!

In view of $\text{tr}(\rho f(\mathbf{x})) = \text{tr}(\rho_0 f(\mathbf{x}))$ (and with obvious definitions) the squared uncertainty $(\Delta x_b)^2$ in a given physical state for a given component of position will equal the usual squared uncertainty $(\Delta_0 x_b)^2$ in the corresponding would-be wave function. Combining this with (22) leads to the replacement of the usual Heisenberg uncertainty relation

$$(\Delta_0 x_a)^2 (\Delta_0 p_b)^2 \geq \frac{1}{4} \delta_{ab}$$

by

$$(\Delta x_a)^2 (\Delta p_b)^2 = (\Delta_0 x_a)^2 ((\Delta_0 p_b)^2 + 2\alpha) \geq \frac{1}{4} \delta_{ab} + 2\alpha (\Delta x_a)^2. \quad (23)$$

We remark that if we were to take this relation to be fundamental then we could recover the lower bound $\Delta p_a \geq 2\alpha$ from it.

It is interesting to notice that, in the past few years, and partly motivated by string theory, many authors have proposed modified uncertainty relations along the schematic lines

$$\Delta x \Delta p \gtrsim \frac{1}{2} + c (\Delta p)^2$$

where c is a constant. Sometimes these are thought of as applying to spacetime coordinates but sometimes to the coordinates of a non-relativistic particle, see e.g. [18]. In the latter case, these resemble (23) – and even more closely, in the case $a = b$, its approximation to first order in α :

$$\Delta x_a \Delta p_a \geq \frac{1}{2} + \alpha (\Delta x_a)^2$$

except that the roles of Δx_a and Δp_a are interchanged, so that, e.g. in [18] and other related references, one deduces a lower bound on Δx rather than on Δp .

4. The experimental indistinguishability from standard quantum mechanics and the corrected pragmatic interpretation

In this section, we elucidate further and prove, the *Position Measurement Theorem* which was outlined in the introduction, according to which any sort of position measurement in the Newtonian approximate theory of [2] based on the hypothesis of [1] will, under the naive pragmatic interpretation, have an outcome indistinguishable from that predicted by standard quantum mechanics. We also amplify on what was said in the introduction about the *corrected pragmatic interpretation* and on the conclusion that, with this interpretation, the Newtonian approximation of [2] is experimentally indistinguishable from standard quantum mechanics. We also discuss in detail in subsection 4.1 how this works out for the class of experiments recently proposed by Roger Penrose which we mentioned in the Introduction.

Consider first our example (introduced in Section 1 where we pointed out that it is approximately described – when $\delta \ll R$ – by the one-dimensional Gaussian model, also introduced there) involving a uniform mass-density bead of mass M and radius R free to slide on a straight wire which terminates at stops a distance $2(R + \delta)$ apart. Calling the direction of the wire ‘the x -axis’ and the position of the centre of the bead when it is half-way between the stops ‘the origin’, the would-be wave function, $\psi(x)$, of the centre-of-mass degree of freedom of the bead in its ground state will, of course, be given by

$$\psi(x) = \delta^{-1/2} \cos(\pi x / 2\delta) \tag{24}$$

and thus according to standard quantum mechanics, its momentum in the x -direction will take one of the two values $\pm \pi / 2\delta$ with equal probabilities and hence the expectation value, $\langle p^2 \rangle$, will be $\pi^2 / 4\delta^2$, whereas, by (15), according to our naive pragmatic interpretation, it will be approximately

$$\langle p^2 \rangle = \pi^2 / (4\delta^2) + 18M^2 / R^2. \tag{25}$$

In standard quantum mechanics, one way to experimentally determine the expectation value of p^2 (or indeed to obtain the full statistical distribution of p values) of the state of such a system might be to rigidly attach a light flat mirror (with diameter much larger than δ) to the bead, perpendicular to the x direction and, after each time resetting the bead to be in the state under investigation, repeatedly to reflect off the mirror a small probe particle (with mass much less than M) whose quantum mechanical wavepacket is each time prepared to be an approximately monochromatic pencil-shaped wavepacket with diameter much bigger than δ , aimed at the mirror along a line at an angle of $-\pi/4$ to the x -axis (see Figure 1), with approximate momentum \mathbf{P} pointing in the direction of the pencil and say $\langle \text{xvii} \rangle$ much larger in magnitude than $\pi/(2\delta)$. If the mirror was fixed rigidly, this pencil would of course turn through a right-angle when it reflects off the mirror but because the mirror is fixed to the bead, its (standard quantum-mechanical) state will be a superposition (or more generally a mixture of superpositions) of wave functions $\psi_n(x) = \delta^{-1/2} \sin(n\pi x/2\delta)$ each of which, in its turn, is of course a superposition of states with momenta $\pm p_n = \pm n\pi/2\delta$. Therefore (see Note $\langle \text{xvii} \rangle$) the probe particle would be predicted to emerge in one of the (approximate) directions $+\pi/4 \pm \sqrt{2}p_n/P$ with a probability depending on the state in the usual way.

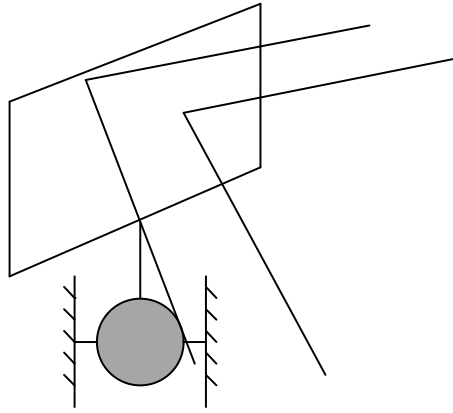


Figure 1. Bead on wire with stops fitted with mirror reflecting a pencil of light

By observing the intensity distribution of probe particles on a suitable screen one could thus infer the probability distribution of the state of the bead and, in particular, deduce the value of $\langle p^2 \rangle$. In particular, if the mirror were in its ground state (24), one would have probability 1/2 of the probe particle emerging in each of the (approximate) directions $+\pi/4 \pm \sqrt{2}p_1/P$ (two spots on the screen of equal intensity) and infer $\langle p^2 \rangle = p_1^2 = \pi^2/(2\delta)^2$ while, in any other state, other directions would be populated (two more widely spaced spots or more than two spots on the screen etc.) and the value inferred for $\langle p^2 \rangle$ would be the appropriate standard quantum mechanical value.

However, we will now argue that, on the Newtonian approximation of [2], according to our naive pragmatic interpretation – applied now to a model closed system consisting

of our bead-with-mirror plus probe-particle system – the result of such an experiment will always be identical with the result predicted by standard quantum mechanics, i.e. the result which, according to standard quantum mechanics, would lead one to infer that the value of $\langle p^2 \rangle$ is the standard quantum mechanical value. Thus, in particular, if we always reset the bead to be in its ground state, the value inferred for $\langle p^2 \rangle$ will be $\pi^2/4\delta^2$ and not the value given in (25).

To see this, it is convenient to adopt a ‘time-dependent’ approach to the reflection of our probe particles from the mirror. Denoting the ground state of the bead by $\psi(x)$ and the would-be ‘in’ wave function for a wave-packet description of the state of one of our probe particles at a time well before it hits the mirror by $\phi_{\text{in}}(\mathbf{y})$, so that the total ‘in’ wave function of the bead-probe-particle system is given by

$$\Phi_{\text{in}}(x, \mathbf{y}) = \psi(x)\phi_{\text{in}}(\mathbf{y}),$$

let us denote by $\Phi_{\text{out}}(x, \mathbf{y})$ the time-evolute, according to the appropriate standard Schrödinger time-evolution, of Φ_{in} at the time of observation of the probe particle on the screen and denote by ρ_{in}^0 the projector $|\Phi_{\text{in}}\rangle\langle\Phi_{\text{in}}|$ and by ρ_{out}^0 the projector $|\Phi_{\text{out}}\rangle\langle\Phi_{\text{out}}|$. Then the Newtonian approximation of [2] tells us that the total physical density operators ρ_{in} and ρ_{out} at the two times will be given by (cf. (7))

$$\rho_{\text{in}}(x, \mathbf{y}; x', \mathbf{y}') = \rho_{\text{in}}^0(x, \mathbf{y}; x', \mathbf{y}') \exp(-D(x, \mathbf{y}; x', \mathbf{y}')) \quad (26)$$

and

$$\rho_{\text{out}}(x, \mathbf{y}; x', \mathbf{y}') = \rho_{\text{out}}^0(x, \mathbf{y}; x', \mathbf{y}') \exp(-D(x, \mathbf{y}; x', \mathbf{y}'))$$

where, we remark, amongst its several properties, the only relevant property of D for the present argument will be that

$$D(x, \mathbf{y}; x, \mathbf{y}) = 0. \quad (27)$$

In the above thought-experiment, the result of any measurement of the statistical distribution of probe particles on the screen will, according to the naive pragmatic interpretation, arise from a formula of the form $\text{tr}(f(\mathbf{y})\rho_{\text{out}})$ where $f(\mathbf{y})$ is the same function of the position of the probe particle that would be taken to represent the relevant observable in standard quantum mechanics. We have

$$\text{tr}(f(\mathbf{y})\rho_{\text{out}}) = \int \int f(\mathbf{y})\rho_{\text{out}}(x, \mathbf{y}; x, \mathbf{y}) dx d^3y$$

But in view of (26) and (27) this is the same thing as

$$\int \int f(\mathbf{y})\rho_{\text{out}}^0(x, \mathbf{y}; x, \mathbf{y}) dx d^3y$$

which is of course the same result which would be predicted for the same measurement in standard quantum mechanics. In particular, the statistical distribution of probe particles observed on the screen when we perform our experiment with the mirror in its ground state will be identical with that predicted according to standard quantum mechanics, which is what we set out to show.

One might of course choose to do a different sort of experiment to determine $\langle p^2 \rangle$ of the state of our mirror – perhaps involving photons as probe particles and making use of optical devices such as interferometers etc. There is then, admittedly, some uncertainty as to what the predictions of the theory in [1] would be because of the relativistic nature of the photon and hence the inapplicability of the Newtonian approximation of [2]. But assuming (we shall return to this below) photons can, for this purpose, be treated as if they were non-relativistic particles (perhaps with a tiny mass) then e.g. an experimental arrangement which results in a detection of which of two optical channels a probe particle or photon goes down or an experimental arrangement which results in the examination of an interference pattern on a screen would, by these arguments, both give the identical result to that predicted by standard quantum mechanics, since both sorts of measurements are types of position measurement.

We remark that similar considerations lead us to conclude that the results of experiments along the lines of those of Folman et al [8], to detect changes in parity of certain quantum mechanical states due to possible deviations from standard quantum mechanical laws, will, on the theory of [2] (and continuing to assume our reasoning continues to apply when some of the particles are photons) be identical to the results predicted by standard quantum mechanics, in spite of the fact that our naive pragmatic interpretation predicts changes in the expectation value of parity as in our formula (18). The reason for this is again that the Folman et al experiment ultimately involves a type of position measurement (again of a photon).

In fact quite generally, for any sort of experiment, as long as the result is obtained by a position measurement of some sort, to the extent that the approximation of [2] is valid, if we apply the theory of [2] to the closed total system consisting of the system of interest together with any relevant probe particles or ‘pointer variables’ etc. and adopt our naive pragmatic interpretation for this closed total system, then an adaptation of the above proof in the case of the bead example shows that the result of our measurement will be identical to the result predicted by standard quantum mechanics. To see this, suppose the total system consists of N bodies in all (protons, electrons etc.) and replace the variables (x, \mathbf{y}) in the above proof by the N positions of all of these. (It may be helpful to think of x as standing for, say, n ‘system’ variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ and \mathbf{y} as standing for $m = N - n$ ‘pointer variables’ $\mathbf{y}_1, \dots, \mathbf{y}_m$ and we shall refer to this way of thinking in the next few parenthetical remarks.) Then the above proof goes through, thinking now of ρ_{in}^0 as the total would-be density operator at the time of preparation of the experiment and ρ_{out}^0 as its Schrödinger time-evolute at the time of the measurement, $f(\mathbf{y})$ as the appropriate position operator (i.e. an appropriate function of all the $\mathbf{y}_1, \dots, \mathbf{y}_m$) and taking the integrations to now be over the $3N$ -dimensional total configuration space. The key point is again that the relevant decoherence exponent (now a function of the $2N$ vector variables $\mathbf{x}_1, \dots, \mathbf{x}_n; \mathbf{y}_1, \dots, \mathbf{y}_m; \mathbf{x}'_1, \dots, \mathbf{x}'_n; \mathbf{y}'_1, \dots, \mathbf{y}'_m$) will vanish on the diagonal (i.e. when $\mathbf{x}_a = \mathbf{x}'_a \quad \forall a \in \{1, \dots, n\}, \mathbf{y}_b = \mathbf{y}'_b \quad \forall b \in \{1, \dots, m\}$). This completes the proof of our general *Position Measurement Theorem* which we referred to in the Introduction.

In our *corrected pragmatic interpretation* we assume that a realistic measurement actually consists of a whole (Heisenberg-von Neumann) chain of measurements – each of which is a measurement in the sense we have been discussing up to now – as explained in the very well-known discussion in Chapter 6 of von Neumann’s book [10]. Thus, for example, in the bead/mirror-probe-particle example with which we opened this section, the dynamics of the bead/mirror-probe-particle system are not the end of the story: the measurement of the position of the probe particle on the screen can in turn be analyzed by including the physics behind the formation of our spots on the screen. Let us suppose, for example that the spots consist of black grains of metallic silver formed when our probe particle hits a silver-iodide-coated screen. Then we would incorporate the screen too in the total quantum system and include suitable terms in the total Hamiltonian to describe the formation of the silver granules. Then one could perform yet another step of this sort to incorporate the physical mechanism by which light is used to measure the position of the silver-granule-spots and then by which the eye records the state of the light etc. etc. However far down such a chain one goes, our above *Position Measurement Theorem* will ensure that one will obtain the same result which is predicted by standard quantum mechanics, provided only that, at the stage of the chain at which one chooses to stop and apply the rule (13), the measurement is a sort of position measurement. We could then simply declare <xviii> that the only realistic measurements are those where one does stop at a point of the von Neumann chain which consists of a sort of position measurement. Alternatively and again following a well-known line of argumentation (cf. e.g. [28]) and cf. also the interesting recent paper [29] of Geoffrey Sewell) one could argue that, whatever may be the cause of the irreversibility, it is, in practice, only at a stage of the chain after something macroscopic and in practice irreversible has happened (such as the formation of our silver granules) that a measurement has really been made <xix> and thereafter it is, in practice, immaterial whether the subsequent stages of the chain are analysed with quantum or with classical physics and, in view of that, it cannot make any difference if, insisting on still analysing quantum mechanically things which might be analysed classically, we decide to choose a stage of the chain in which the measurement is a sort of position measurement – e.g. a pointer reading on a dial.

Our *corrected pragmatic interpretation* consists in adopting either of these points of view; it doesn’t matter which. All that matters is that, for one reason or another, we take the view that a realistic measurement consists of a sort of position measurement at some point of the Heisenberg-von Neumann chain. With our above *Position Measurement Theorem* we then immediately conclude (as already stated in the introduction)

On our corrected pragmatic interpretation, the experimental predictions of the Newtonian approximation of [2] are identical with those of standard quantum mechanics!

(A)

Returning momentarily to our bead/mirror-probe-particle example, we remark that, actually, if the probe particle is sufficiently light (but we continue to use the non-

relativistic theory of [2]) then $D(x, \mathbf{y}; x', \mathbf{y}')$ will be well-approximated by a function, say d , of x and x' only, so that

$$\rho_{\text{out}}(x, \mathbf{y}; x', \mathbf{y}') \simeq \rho_{\text{out}}^0(x, \mathbf{y}; x', \mathbf{y}') \exp(-d(x; x')).$$

Thus, for any observable A which refers to the probe particle so that, formally, it can be represented by a kernel $A(\mathbf{y}, \mathbf{y}')$ which is only a function of \mathbf{y} and \mathbf{y}' , we will have

$$\text{tr}(A\rho_{\text{out}}) = \int \int \int A(\mathbf{y}, \mathbf{y}') \rho_{\text{out}}(x, \mathbf{y}'; x, \mathbf{y}) dx d^3y d^3y'$$

which by (4) is approximately equal to

$$\int \int \int A(\mathbf{y}, \mathbf{y}') \rho_{\text{out}}(x, \mathbf{y}'; x, \mathbf{y}) dx d^3y d^3y'$$

which, in view of $d(x, x) = 0$, is equal to

$$\int \int \int A(\mathbf{y}, \mathbf{y}') \rho_{\text{out}}^0(x, \mathbf{y}'; x, \mathbf{y}) dx d^3y d^3y'$$

i.e. to

$$= \text{tr}(A\rho_{\text{out}}^0).$$

This result will obviously generalize so as to conclude that, on our naive pragmatic interpretation, the results of *any* sort of measurement on a probe particle at some stage of a Heisenberg-von Neumann chain of measurement, will, if the probe particle is sufficiently light, be *approximately* the same as the results predicted by standard quantum mechanics.

However, we wouldn't assign as much importance to this result as to our *Position Measurement Theorem* because it is only approximate and also because, in some given measurement chain, it may only apply e.g. at a stage of the measurement chain before classicality and irreversibility have set in. In such a case, we would still rather rely on our Position Measurement Theorem, applied at a later stage of the measurement chain where the measurement is, e.g. on a macroscopic pointer (which may well not be so light that one could make the above approximation) in order to justify the statement (A).

As we have mentioned, all our results assume that the entire chain of measurements involves only non-relativistic particles, whereas, in practice (an example is the Penrose experiment we shall discuss in detail below) photons are often involved at some stage(s) of the chain. This is a gap which needs to be filled. However, we would expect that, provided (in the lab frame) the relevant photon states are not too energetic and provided they are sufficiently far from being on the verge of black-hole collapse then they may be treated as if they are non-relativistic particles with a tiny mass. In particular, we shall assume this to be the case for the photons in the Penrose experiment which we discuss next.

Of particular interest to us are interferometer-like experiments along the lines of those suggested by Penrose [21], [22] to detect the 'collapse of the wavepacket' of certain 'Schrödinger Cat-like' macroscopic quantum superpositions as predicted by

certain theoretical proposals for modifications of standard quantum mechanics, notably collapse-model proposals such as that of GRW [13] and the proposal of Penrose himself [20] concerning his gravity-induced ‘quantum state reduction’. In one of its simpler versions <xx>, the setup proposed by Penrose is reproduced here in Figure 2.

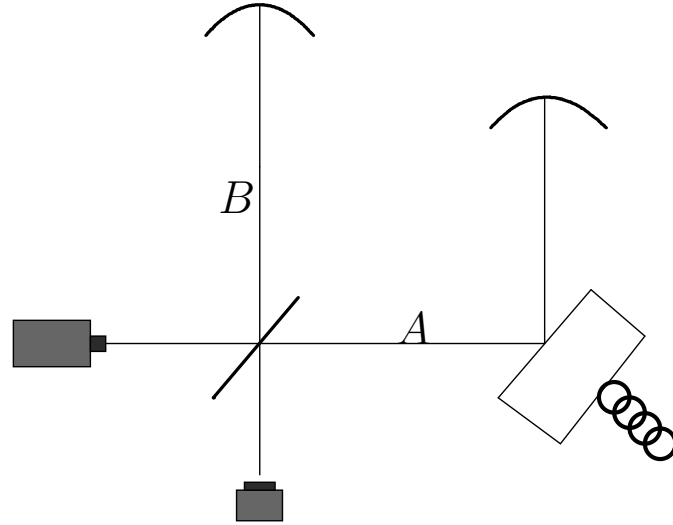


Figure 2. Diagram of the Penrose experiment

As described in [21], a photon, emitted from a source, is passed through a beam-splitter and emerges in a superposition of two components with paths at an angle to one another, one, say ‘component *B*’, of which is reflected directly back towards the beam splitter by a fixed mirror, whereas the other, say ‘component *A*’, while eventually reflected back by a second fixed mirror, also reflects off a movable mirror of macroscopic dimensions and mass, both on its way towards and on its way back from that fixed mirror. The movable mirror is attached to an oscillator and (in a classical description) initially at rest. The mass of the movable mirror, the spring constant of the oscillator, the frequency of the photon and the spacing of the movable mirror and the second fixed mirror are tuned so that (staying with a classical description)

in component ‘A’ of the superposition (in which the photon hits the movable mirror) its first encounter will set the movable mirror into motion, while its second encounter will occur at the movable mirror’s initial location and bring it back to rest there – the movable mirror swinging to and from a macroscopically different location in between.

(α)

The positions of all the mirrors are also tuned so that, according to a standard quantum mechanical analysis,

both components of the photon would arrive back at the beam splitter at the same time and with relative phases such that they reassemble into a state which travels back towards

the photon source, with no photons detected at the detector (see Figure 2) placed at the appropriate angle to that direction to catch any photons which might have emerged from the beam-splitter in its alternative direction.

This standard quantum mechanical prediction depends however on the laws of quantum mechanics applying even though, at an intermediate stage of the experiment (i.e. between the two reflections of component *A* of the photon with the movable mirror) the total quantum state of the system is a superposition involving two macroscopically different positions of a macroscopic object, i.e. of the movable mirror. If a ‘wave-function collapse’ of this superposition occurs, e.g. as predicted by Penrose’s 1996 proposal [20], then one would predict that, in consequence, the coherence of the superposition of the two components of the photon beam as they reconverge on the beam-splitter will be lost and therefore that photons will arrive at the source and at the detector with equal probabilities.

In this way, the Penrose experiment should be able to distinguish between standard quantum mechanics and the gravitational-induced quantum state reduction proposal of Penrose 1996 [20] (or between standard quantum mechanics and one or other collapse model).

One might at first sight expect that Kay’s 1998 theory (i.e. the Newtonian approximate theory [2] based on Kay’s hypothesis [1]) would also predict a similar non-standard result for this Penrose experiment (for suitably large crystals etc.). After all, as we have seen, for suitably large crystals etc., that theory too predicts the decoherence of superpositions involving macroscopically different configurations. However, our general result above tells us (assuming, as we have discussed above, that the result still applies when the probe particles whose position is being measured are photons) that this can’t be. Just as for any other experiment which ends in a position measurement, the prediction must be identical with the result predicted by standard quantum mechanics. This difference between the predictions of Penrose’s 1996 proposal [20] and of Kay’s 1998 theory [2] for the Penrose experiment may be attributed to the following difference: In Penrose’s work [20] it is assumed that once a macroscopic system enters a sufficiently macroscopic superposition (in the present case involving the movable mirror in two macroscopically different locations after the first reflection off it by Component *A* of the photon) then that superposition will collapse *and remain collapsed irrespective of its future dynamics*. (The same property holds in most collapse models.) In [2] on the other hand, the macroscopic superposition formed after the first reflection of component *A* of the photon at the movable mirror will decohere but then, at the second reflection *recohere* (but we emphasize that this is a feature of the Newtonian approximate theory [2], not of the full hypothesis of [1] – see the remarks at the end of the following subsection and in Section 7). We explain this point more fully in the following subsection.

4.1. Further discussion of the Penrose experiment

In a mathematical description of the above standard quantum mechanical analysis, the time-evolving Schrödinger wave function of the photon/movable-mirror-cum-oscillator system may be regarded as an element of a total matter Hilbert space

$$\mathcal{H}_{\text{matter}} = \mathcal{H}_{\text{photon}} \otimes \mathcal{H}_{\text{mirror}}.$$

Suppose, in a wave packet description of the photon, the photon state at a time immediately after the photon emerges from the beam splitter takes the form $\frac{1}{\sqrt{2}}(\gamma_A(0) + \gamma_B(0)) \in \mathcal{H}_{\text{photon}}$ where $\gamma_A(0)$, $\gamma_B(0)$ represent the A , B components respectively of the photon superposition at that time, and suppose that the initial state, $\mu(0) \in \mathcal{H}_{\text{mirror}}$, of the movable-mirror-cum-oscillator is its ground state, then we might interpret our statement (α) by the statement that $\gamma_A(0) \otimes \mu(0) \in \mathcal{H}_{\text{matter}}$ will have evolved, by an intermediate time, t_m (the suffix ‘m’ here stands for ‘mid’) shortly after the first reflection of the photon off the movable mirror, say, to a state $\gamma_A(t_m) \otimes \mu(t_m)$ where $\mu(t_m)$, being a quantum description of the movable mirror in its macroscopically different location, will approximately satisfy $\langle \mu(t_m) | \mu(0) \rangle = 0$, but that by the later time, t_f (‘f’ here standing for ‘final’), when the photon has undergone its second reflection off the movable mirror and is about to re-enter the beam splitter in the return direction, it will have further evolved to $\gamma_A(t_f) \otimes \mu(t_f)$ where, as the quantum counterpart to the classical statement that the movable mirror will have returned to its rest position, $\mu(t_f)$ will be taken to be equal to (a phase times) the ground state $\mu(0)$. On the other hand, Component B of the photon doesn’t interact with the movable mirror at all, so the vector $\gamma_B(0) \otimes \mu(0) \in \mathcal{H}_{\text{total}}$ will evolve at any time into (a phase times) $\gamma_B(t) \otimes \mu(0)$ say.

Putting all this together, standard quantum mechanics therefore would predict that the total state of the photon/movable-mirror-cum-oscillator system at time t_f just before (both components of) the photon re-enter(s) the beam splitter will be <xxii>

$$\Psi_{\text{matter}}^{\text{standard}}(t_f) = \frac{1}{\sqrt{2}}(\gamma_A(t_f) + \gamma_B(t_f)) \otimes \mu(0) \quad (28)$$

so that the partial state of the photon itself (i.e. after tracing $|\Psi_{\text{matter}}^{\text{standard}}\rangle\langle\Psi_{\text{matter}}^{\text{standard}}|$ over $\mathcal{H}_{\text{mirror}}$) will be the pure state with density operator

$$\rho_{\text{photon}}^{\text{standard}}(t_f) = \frac{1}{2}|(\gamma_A(t_f) + \gamma_B(t_f))\rangle\langle(\gamma_A(t_f) + \gamma_B(t_f))| \quad (29)$$

and we assume that this is the photon state which passes entirely through the beam splitter in the direction of the source, leading to zero detection probability at the detector.

Now, what we understand to be envisaged by Penrose [21], [22] <xxiii>, is that, on his 1996 proposal [20], the initial total density operator of the photon/movable-mirror-cum-oscillator system

$$\rho(0)_{\text{matter}}^{\text{Penrose}} = \frac{1}{2}|(\gamma_A(0) + \gamma_B(0))\rangle\langle(\gamma_A(0) + \gamma_B(0))| \otimes |\mu(0)\rangle\langle\mu(0)|$$

will (because of gravitational-induced quantum state reduction of the macroscopic movable-mirror-cum-oscillator superposition according to his proposal) evolve by time t_m into a total density operator close to

$$\rho_{\text{matter}}^{\text{Penrose}}(t_m) = \frac{1}{2}|\gamma_A(t_m)\rangle\langle\gamma_A(t_m)|\otimes|\mu(t_m)\rangle\langle\mu(t_m)| + \frac{1}{2}|\gamma_B(t_m)\rangle\langle\gamma_B(t_m)|\otimes|\mu(0)\rangle\langle\mu(0)| \quad (30)$$

and, because it is (at least implicitly) assumed that a reduced wave function cannot ‘dereduce’ it is assumed that this will evolve by t_f into

$$\rho_{\text{matter}}^{\text{Penrose}}(t_f) = \frac{1}{2}|\gamma_A(t_f)\rangle\langle\gamma_A(t_f)|\otimes|\mu(0)\rangle\langle\mu(0)| + \frac{1}{2}|\gamma_B(t_f)\rangle\langle\gamma_B(t_f)|\otimes|\mu(0)\rangle\langle\mu(0)|.$$

and therefore that the partial state of the photon itself as it re-enters the beam-splitter will be

$$\rho_{\text{photon}}^{\text{Penrose}}(t_m) = \frac{1}{2}|\gamma_A(t_f)\rangle\langle\gamma_A(t_f)| + \frac{1}{2}|\gamma_B(t_f)\rangle\langle\gamma_B(t_f)|$$

which is expected to predict equal probabilities of detecting a photon at the source and at the detector.

On the other hand, the theory of [1], under the approximation of [2] would (see Notes <vi> and <viii>) require the replacement of the initial density operator $|\Psi_{\text{matter}}^{\text{standard}}\rangle\langle\Psi_{\text{matter}}^{\text{standard}}|$, $\Psi_{\text{matter}}^{\text{standard}}$ as in (28) on $\mathcal{H}_{\text{matter}} = \mathcal{H}_{\text{photon}} \otimes \mathcal{H}_{\text{mirror}}$ by the partial trace, $\rho_{\text{matter}}^{\text{Kay}}$ of $|\Phi_{\text{total}}^{\text{Kay}}\rangle\langle\Phi_{\text{total}}^{\text{Kay}}|$ over $\mathcal{H}_{\text{gravity}}$ where $\Phi_{\text{total}}^{\text{Kay}}$ is a suitable replacement for $\Psi_{\text{matter}}^{\text{standard}}$ on $\mathcal{H}_{\text{total}} = \mathcal{H}_{\text{matter}} \otimes \mathcal{H}_{\text{gravity}} (= \mathcal{H}_{\text{photon}} \otimes \mathcal{H}_{\text{mirror}} \otimes \mathcal{H}_{\text{gravity}})$. In the formalism of [2] (cf. Notes <vi> and <viii>) one would take (assuming as above that photons may be treated for this purpose as non-relativistic particles) $\Psi_{\text{total}}^{\text{Kay}}(t)$ to be given by

$$\Psi_{\text{total}}^{\text{Kay}}(t) = \frac{1}{\sqrt{2}}(\gamma_A(t) \otimes \mu(t) \otimes g(t) + \gamma_B(t) \otimes \mu(0) \otimes g(0))$$

where $g(t)$ is the (non-radiative) quantum counterpart to the classical Newtonian gravitational field of a (static) mirror at the position of the movable mirror at time t when the photon is in Component ‘A’ of its superposition. $\rho_{\text{matter}}^{\text{Kay}}(t)$ will then clearly take the following values at times 0, t_m and t_f : At time zero, it will coincide with $\rho_{\text{matter}}^{\text{standard}}(0)$; at time t_m it will coincide with $\rho_{\text{matter}}^{\text{Penrose}}(t_m)$, but at time t_f , it will again coincide with $\rho_{\text{matter}}^{\text{standard}}(t_f) = |\Psi_{\text{matter}}^{\text{standard}}(t_f)\rangle\langle\Psi_{\text{matter}}^{\text{standard}}(t_f)|$ (see (28)). In other words, between times t_m and t_f , $\rho_{\text{matter}}^{\text{Kay}}$ will ‘recohere’ and in fact, as we know it must, will therefore result in identical experimental results to those predicted by standard quantum mechanics <xxiv>.

The reason for this recoherence is (cf. Note <viii> and also the last paragraph in Section 6) because the gravitational state-vector $g(t)$ depends only on the state of the movable mirror at the time t and not on its history and therefore, since (reverting to the classical description which is relevant here) the movable mirror returns to its initial position at time t_f , we will have $g(t_f) = g(0)$ <xxv>. Of course, in a fully relativistic treatment, this is expected no longer to be the case and in particular, if the motion of the movable mirror between times 0 and t_f causes the emission of one or more ‘gravitons’

then one would expect $\langle g(t_f)|g(0)\rangle = 0$ (approximately) in consequence of which the partial density operator $\rho(t_f)$ of the photon at time t_f (and hence the experimental prediction) will coincide with $\rho_{\text{matter}}^{\text{Penrose}}(t_f)$. However, an order of magnitude estimate shows that, with typical experimental values for the mass and size and frequency of the movable mirror, the amplitude to emit a graviton will be tiny on time-scales relevant to the experiment <xxvi>.

5. A tentative interpretation in terms of events

As discussed in the Introduction around the passage labelled (B) and discussed further in Notes <xiv>, <xv>, <xvi>, given any candidate for a fundamental alternative theory to standard quantum mechanics which is couched in terms of a time-evolving density operator, $\rho(t)$, it would seem natural to attempt an interpretation in terms of ‘events’ which ‘happen’, by identifying the set of events which can happen at a given time t with the set of spectral subspaces of $\rho(t)$ and identifying the probability with which a given such event happens with the associated eigenvalue, λ , multiplied by its multiplicity, m . Thus, in particular, it is interesting to explore the prospects for such an interpretation in the case of the hypothesis of [1] and one might hope that it would have a more fundamental status than either of our ‘naive’ or ‘corrected’ pragmatic interpretations (i.e. of the Newtonian approximate theory of [2]). From now on (and in Note <xxvii>) we shall refer to it as our ‘events’ interpretation. To make a small beginning in this, we shall begin to explore this question here in the context of the Newtonian approximation of [2]. The caveats mentioned at the beginning of Section 1.2 should of course continue to be borne in mind. (They will be discussed further Section 7).

As we have already indicated in the Introduction and in the notes mentioned above, in the case of a would-be wave function for the centre of mass of a single ball in a Schrödinger Cat-like state

$$\psi = c_1\psi_1 + c_2\psi_2$$

where ψ_1 and ψ_2 are sharply localized around two different locations, provided the mass of the ball is much larger than the Planck mass and the distance between the two locations a big enough fraction of the ball radius, the physical density operator will be well-approximated by (12)

$$\rho = |c_1|^2|\psi_1\rangle\langle\psi_1| + |c_2|^2|\psi_2\rangle\langle\psi_2|.$$

Our events interpretation then tells us that (except when $|c_1|^2 = 1/2 = |c_2|^2$) two events are possible, one corresponding to the subspace spanned by ψ_1 , and one corresponding to the subspace spanned by ψ_2 . Further, these events will occur with probabilities $|c_1|^2$, $|c_2|^2$ respectively. This is arguably just what one would hope for from a resolution to the Schrödinger Cat puzzle – see Note <xiii> for further discussion.

In the remainder of this section, we shall mainly discuss how our events interpretation works out for a submodel of the one-dimensional Gaussian model which was introduced in Section 1, postponing discussion of more general models to Section

7. We recall from Section 1 that in this model the relationship between the would-be wave function, ψ (on the line) and the physical density operator, ρ , is given by

$$\rho(x, x') = \psi(x)\psi^*(x')e^{-\kappa(x-x')^2} \quad (31)$$

and that the model is expected to give a good approximation to the physical density operator for the centre-of-mass motion of a uniform-mass-density bead of mass M and radius R constrained to move along a straight wire provided M is around or bigger than the Planck mass, or, in the case the wire terminates at hard stops a distance $2(R + \delta)$ apart, for any mass provided $\delta \ll R$.

We shall restrict ourselves further to would-be wave functions ψ in this model which are either even or odd functions of x . As we show in the Appendix, using perturbation-theoretic methods, it is very easy to explicitly diagonalize any ρ in (31) for a ψ which satisfies this further restriction, to first order in κ , and one finds, to this order, that there are only two non-zero eigenvalues (each non-degenerate),

$$\lambda_1 = 1 - 2\kappa\langle x^2 \rangle, \quad \lambda_2 = 2\kappa\langle x^2 \rangle \quad (32)$$

with corresponding eigenfunctions

$$\phi_1 = c_1((1 + \kappa\langle x^2 \rangle)\psi - \kappa x^2 \psi), \quad \phi_2 = c_2 x \psi. \quad (33)$$

where c_1 and c_2 are normalization constants. Here and below, $\langle x^2 \rangle$ stands for the expectation value $\langle \psi | x^2 | \psi \rangle$ of x^2 in the would-be state ψ .

As long as $\kappa\langle x^2 \rangle$ is very much less than 1,

$$\kappa\langle x^2 \rangle \quad (= \kappa\langle \psi | x^2 | \psi \rangle) \quad \ll 1, \quad (34)$$

the eigenvalues (32) will be close to 1 and zero and therefore one expects this first-order perturbation theory result to give a good approximation to the true spectrum.

We conclude that, if the would-be wave function, ψ , satisfies $\langle \psi | x^2 | \psi \rangle \ll 1$ and is either an even function or an odd function of x then, to a good approximation, the possible events, according to our events interpretation, will be the one-dimensional subspaces spanned by ϕ_1 and ϕ_2 in (33) and their respective probabilities will be λ_1 and λ_2 in (32). To illustrate this, we have sketched, in Figure 3, ϕ_1 and ϕ_2 when the would-be wave function is the ground state (24)

$$\psi(x) = \delta^{-1/2} \cos(\pi x / 2\delta)$$

of our bead-on-a-wire example (see Section 1 after equation (6) and equation (24) in Section 4 and below) with stops at $\pm(R + \delta)$.

In fact, as explained in Section 1, as long as $\delta \ll R$ then the Gaussian model will give a good approximation for the physical density operator, ρ , of this bead-on-a-wire-with-stops example whatever the mass, while, in view of (35) and, recalling that $\kappa = 9M^2/R^2$, we expect the above first-order perturbation theory results to be applicable provided $\delta \ll R/3M$ which is of course not an additional restriction as long as M is around or smaller than the Planck mass.

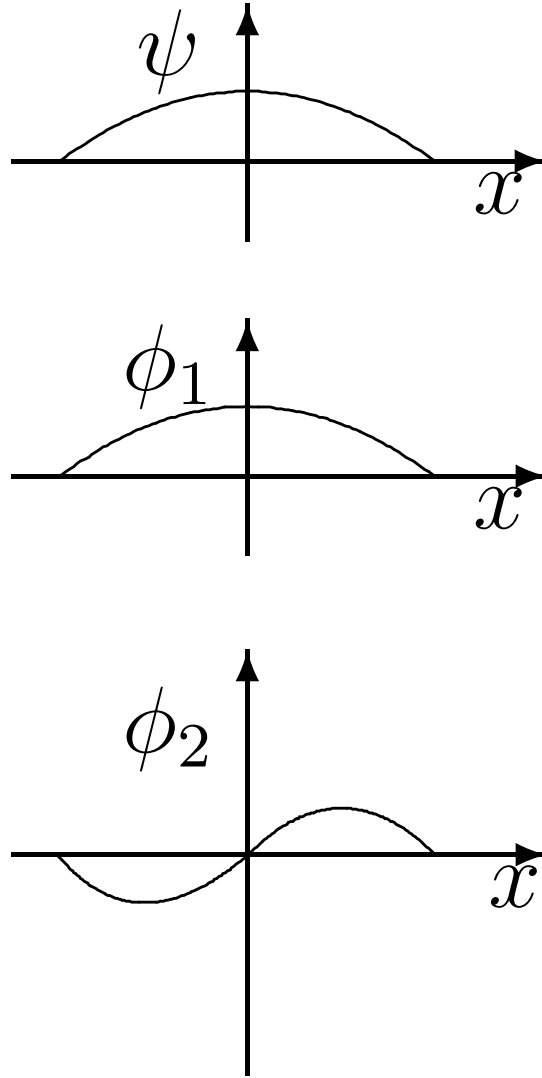


Figure 3. A typical wave function for the one-dimensional Gaussian model and its two ‘events’

On the other hand, we note that for the application to our bead moving along an infinite straight wire, then the condition (34) amounts, in view of $\kappa = 9M^2/R^2$, to the condition

$$\langle \psi | x^2 | \psi \rangle^{\frac{1}{2}} \ll R/3M. \quad (35)$$

so, since, as we saw in Section 1, we need to assume M to be around or greater than 1 for the Gaussian model to be a good approximation at all, we can trust our above first-order perturbation theory result only when the width of the would-be wave function for the centre-of-mass of the bead is very much less than the bead’s radius.

In any case (i.e. whether or not the one-dimensional Gaussian model is a good

approximation for some bead-on-a-wire system) it is interesting to notice (see Figure 3) that (to first order in κ) if we take, say, an even would-be wave function ψ , then ϕ_1 is even and approximately resembles ψ and occurs with a probability, λ_1 , close to 1, but there is a small but non-zero probability, λ_2 , for ϕ_2 to occur and this is a quite different function from (in fact it is, of course, orthogonal to) ψ and it is odd! Moreover, the expectation value of parity $\langle \mathcal{P} \rangle$, by which we now mean the ordinary appropriately weighted sum of probabilities $(+1)\lambda_1 + (-1)\lambda_2$ is then, by (32), given by

$$\langle \mathcal{P} \rangle = 1 - 4\kappa \langle \psi | x^2 \psi \rangle. \quad (36)$$

and coincides with the formula (20) for the expectation value of parity (albeit defined there, differently, to be $\text{tr}(\rho \mathcal{P})$!) obtained under the naive pragmatic interpretation in Section 2. (And one can similarly reproduce (21) when ψ is odd.) The coincidence is of course explained by the fact that, for a would-be wave function, ψ , which is either even or odd, the conventional Parity operator \mathcal{P} is simultaneously diagonalizable with ρ and has eigenvalue $+1$ on ϕ_1 and -1 on ϕ_2 . One can also show, using the results in Subsection 9.2 in the Appendix, that similar coincidences obtain, for the analogous pair of different possible meanings for the phrase ‘expectation value of parity’ in the case of the three-dimensional situation discussed in the sentence containing equation 18 and the subsequent sentence.

6. Entropy and its time dependence

As soon as one knows how to diagonalize ρ , one can calculate its (von Neumann) entropy, $S = -\text{tr}(\rho \ln \rho)$, and other entropy-like measures of its amount of decoherence – in particular $S_1 = 1 - \text{tr} \rho^2$ (see e.g. [24]) which, while less significant physically, is often more tractable mathematically.

Staying with the one-dimensional Gaussian model and working to lowest order in κ as explained in Section 5 and the Appendix, one finds, for a given would-be wave function ψ ,

$$S = -2\kappa \langle x^2 \rangle \ln(2e\kappa \langle x^2 \rangle),$$

$$S_1 = 4\kappa \langle x^2 \rangle$$

where, as in the previous section, $\langle x^2 \rangle$ stands for $\langle \psi | x^2 \psi \rangle$. These are expected to be good approximations as long as $\kappa \langle x^2 \rangle$ is very much less than 1.

If the would-be wave function ψ evolves in time according to some would-be Hamiltonian H , as discussed in the Introduction around equation (8), so that the physical density operator evolves according to the master equation (10)

$$\dot{\rho} = -i[H, \rho] - \frac{2\kappa}{M}[x, [p, \rho]],$$

then, clearly, the time-evolution of the approximate expressions for S and S_1 above is controlled by the time-function $t \mapsto \langle \psi(t) | x^2 \psi(t) \rangle$. (Below, we call this $\langle x^2(t) \rangle$.) If this increases, our approximate S_1 (which is just 4κ times this quantity) will

obviously increase and so will our approximate S provided $\kappa\langle x^2 \rangle < 1/e^2$. (But we are anyway assuming that $\kappa\langle x^2 \rangle \ll 1$.) If $\psi(t)$ evolves according to the free Hamiltonian $H = p^2/2M$, we have the easily derived and well-known ‘spreading-of-the-wavepacket’ result that $\langle x^2(t) \rangle$ will take the quadratic form

$$\langle x^2(t) \rangle = A + (2E/M)(t - t_{\min})^2 \quad (37)$$

where E is the (conserved) expectation value $\langle \psi(t) | H \psi(t) \rangle$ of the energy in the state $\psi(t)$. The feature of (37) to which we wish to draw attention here is that $\langle x^2(t) \rangle$ has a single minimum (with value A) – occurring at t_{\min} . In consequence of this, substituting (37) into either of our above approximate formulae for S or S_1 gives rise to a ‘two-sided’ entropy-increase result in the sense that, if the time t_{\min} is taken as an appropriate ‘zero time’ then the entropy will only increase on either side of that (in other words, it will decrease monotonically before the zero time and increase monotonically after it) at least until $\kappa\langle x^2(t) \rangle$ ceases to be very much less than 1, at which point our approximations and assumptions will break down.

This two-sided entropy increase result would be applicable, for example, to the centre-of-mass degree of freedom of our bead of mass M (see Section 5) bigger than or around the Planck mass, free to slide along an infinite wire and with an initial wave function at t_{\min} such that A in (37) is very much less than $R^2/9M^2$, where R is the bead radius, as long as the times considered are sufficiently short for the wave-packet not to have spread so much that the condition (35) ceases to hold. To give a quantitative example, consider a Planck mass bead ($\approx 2 \times 10^{-5}$ g) with radius $R = 10^{-2}$ cm and a state in which $\langle x^2(t_{\min}) \rangle = A = 10^{-9}$ cm². Then $S_{1\min} (= 4\kappa A = 4 \times \frac{9M^2}{R^2} A) \approx 4 \times 10^{-4}$. Moreover, if we assume that we can trust (32) (and hence our formula for S_1) as long as $\langle x^2 \rangle < 10^{-7}$ cm², say, so that $\kappa\langle x^2 \rangle < 10^{-2}$, then the time, t , at which $\langle x^2(t) \rangle$ attains this value, can be estimated (by setting $\frac{9M^2}{R^2} \frac{2E}{M} t^2 = 10^{-2}$ and taking $E = \frac{p^2}{2M}$ with p^2 by an uncertainty principle estimate $= \frac{1}{4A}$) to be, within an order of magnitude or so, around 10^9 years! ... by which time, S_1 will have increased to around 4×10^{-2} !

It is interesting to compare and contrast this sort of (two-sided) entropy increase result (for either S or S_1) with the sort of entropy increase results which one can obtain for the BLP master equation [27]

$$\dot{\rho} = -i[H, \rho] - c[x, [x, \rho]] \quad (38)$$

(Here H is some choice of Hamiltonian and c a positive constant.) which may be regarded as a prototype example of a master equation of GKS/L form. (See Note <ix>.)

For this BLP model, it is easy to show that, irrespective of the Hamiltonian H , the quantity S_1 will, for positive times, increase monotonically in time. To show this, begin by noticing that it follows immediately from the definition $S_1 = 1 - \text{tr}(\rho^2)$ that the time derivative, $\dot{S}_1 = -2 \text{tr} \dot{\rho} \rho = -2 \text{tr}(-i[H, \rho] \rho) + 2c \text{tr}([x, [x, \rho]] \rho)$.

The first term here vanishes by the cyclic invariance of the trace. The second term equals $-2c \text{tr}([x, \rho]^2) = +2c \text{tr}([x \rho]^\dagger [x \rho])$ which is ≥ 0 . (End of proof.)

This situation differs in at least two important respects from that we saw above for our Gaussian model, which, we recall, satisfies the master equation (10) when H is

taken to be the free Hamiltonian $H = p^2/2M$. To explain the first difference, we first need to notice that the BLP master equation (38) is itself of a quite different nature to (10) in that (see Note <ix>) it holds unrestrictedly for any ρ but only for positive times, whereas (10) is only meaningful for ρ for which $\rho(x, x')$ takes the special form (3) but then holds both for positive and negative times. (See Note <x>.) This difference is of course reflected in the different sorts of entropy increase results which are appropriate, with our two-sided result for (10) and the above, more familiar, one-sided result for (38).

Second, the above ‘one-sided’ entropy-increase result for the BLP master equation (38) holds *irrespective* of the choice of Hamiltonian H in (38). In contrast, the two-sided entropy-increase result for our master equation (10) which we obtained above in the case that H in (10) is the free Hamiltonian $H = p^2/2M$, depended on the form of the Hamiltonian. Indeed, suppose one were to take, instead, for example, the Harmonic oscillator Hamiltonian $H = p^2/2M + kx^2/2$ (describing, say, our bead-on-a-wire when the bead is connected to the origin with a spring with spring constant k) then it is obvious that, since the would-be wave function $\psi(t)$ will be periodic in time (with period $2\pi(M/k)^{1/2}$) then so will S_1 and S – these being functionals of ψ . This second difference can be traced (see Notes <ix> and <x>) to the difference that master equations of GKS/L form such as BLP arise from tracing over ‘radiation’ type modes which ‘fly away’ whereas our master equation (10) arises (cf. the last paragraph of Subsection 4.1 and Notes <viii> and <xxv>) from tracing over modes which (in our bead-on-wire interpretation) are dragged around by the would-be wave function of our bead.

The relevance of the two-sided entropy result obtained in this Section for our Gaussian model to the general understanding of the Second Law entailed by the hypothesis of [1] will be discussed in the next (Discussion) section.

7. Discussion

Our main purpose has been to explore further the non-relativistic approximate theory of [2] which, as argued in [2] and explained in the Introduction, is expected to be the non-relativistic limit of any full theory of quantum gravity consistent with the hypothesis of [1] which, in its turn, as argued in [1] (see also Note <v>) offers a natural resolution to the several puzzles and problems mentioned at the start of the Introduction and in Note <i>.

As we have explained in the Introduction, this approximate theory amounts, in the most general form in which we have stated it (see Equation (7), Note <vii>, and the parenthetical remark after Equation (8)) to the following modification of standard (many-body) non-relativistic quantum mechanics:

One has an ‘underlying’ many-body wave function $\Psi(t; \mathbf{x}_1, \dots, \mathbf{x}_N)$ which evolves in time *exactly* as in standard non-relativistic quantum mechanics for the usual many-body Hamiltonian that one would adopt for the problem of interest. However, while one would normally take the density operator of the system at time t to be given by the

(pure) projector onto Ψ , which in its position-space representation is given by

$$\rho_0(t; \mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N) = \Psi(t; \mathbf{x}_1, \dots, \mathbf{x}_N) \Psi(t; \mathbf{x}'_1, \dots, \mathbf{x}'_N)^* \quad (39)$$

one declares that the *physically relevant* density operator ρ is given, in its position-space representation by

$$\rho(t; \mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N) = \rho_0(t; \mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N) e^{-D(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N)}, \quad (40)$$

where the *decoherence exponent* $D(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N)$ is a specific function on configuration space (see Note <vii>) which depends both on the mass of each body and also on a suitable radius (see the Introduction, Note <vii> and [3]) which is ascribed to each particle and which satisfies the general properties stated after Equation (7) – the most important of which for the present discussion are

(a) D tends to be large for pairs of configurations $(\mathbf{x}_1, \dots, \mathbf{x}_N)$, $(\mathbf{x}'_1, \dots, \mathbf{x}'_N)$ which (in the terminology of [2]) involve ‘mass relocations’ around or bigger than the Planck mass. As discussed further in [2], for a given pair of configurations, the *mass relocation* involved may be defined to be the minimum amount of mass that has to be moved in order to convert the unprimed configuration into the primed configuration (or vice versa).

(b) D vanishes on the diagonal, i.e.

$$D(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}_1, \dots, \mathbf{x}_N) = 0$$

As was discussed in the Introduction and in [2], Property (a) leads to the suppression of macroscopic (‘Schrödinger Cat-like’) superpositions while property (b) is what lies behind our *Position Measurement Theorem* which, on our *corrected pragmatic interpretation* (see Section 4) leads us to conclude that the experimental predictions of our non-relativistic approximate theory are identical to those of standard quantum mechanics. (As we cautioned at the beginning of Subsection 1.2 this conclusion is also predicated on being able to treat the relevant small non-relativistic systems as closed for the purposes of this discussion of measurement. We shall provisionally assume that it is and give arguments which confirm this towards the end of this section.)

As we remarked in the Introduction (see before the paragraph labelled (A) in Section 1.2) the latter conclusion might seem surprising. After all, for large macroscopic systems, the difference between the mathematical quantities ρ and ρ_0 can be huge by any measure. On the other hand, the conclusions of [2] that quantum gravitational effects could have any sort of big effect on every-day laboratory sized quantum systems were themselves surprising. So it is perhaps reassuring that we have now arrived at the conclusion that (in the non-relativistic regime) they don’t have any *measurable* effect after all. The way in which they manage not to have a measurable effect, however, is rather subtle and despite the experimental indistinguishability, the theory of [2] possesses properties and prospects quite different from those of standard quantum mechanics as we now discuss.

First and foremost, the theory of [2] achieves the goal of replacing standard quantum mechanics by a modified theory in which macroscopic (‘Schrödinger Cat-like’) superpositions are absent. As far as the achievement of this goal is concerned, it would,

indeed, seem to be an attractive rival e.g. to the ‘collapse models’ such as that of GRW [13] and others, which achieve the same goal at the expense of certain ad hoc modifications of quantum mechanics (and which also entail certain small modifications to the experimental consequences).

It is interesting to notice that, in the light of the present work, one realizes that it would have been possible to achieve the same goal simply by positing that the usual pure density operator (39) of standard quantum mechanics should be replaced by a physical density operator given by (40) with D any (ad hoc) choice of function on $Configuration\ space \times Configuration\ space$ (perhaps quite unrelated to quantum gravity) designed to have the properties (a) and (b) above (and the other general properties stated after Equation (7)). For, once one posits any such replacement, one will be able to prove a *Position Measurement Theorem* and to propose a *corrected pragmatic interpretation* etc. To the best of our knowledge, however, no such proposal had hitherto been made. In any case, the proposal with our approximate theory of [2] would (in the absence of any current experimental guidance) seem to be more satisfactory than that with any such ad hoc choice of D and also more satisfactory than the resolution achieved with any ad hoc collapse model, not only because it is less ad hoc, but also insofar as it is part and parcel of the hypothesis of [1] which offers a much more wide-ranging collection of simultaneous resolutions to several other puzzles as explained at the outset of the Introduction and in Notes <i>, <iii>, <iv>, <v> and <xii>.

The fact that our non-relativistic approximate theory manages to resolve the Schrödinger Cat puzzle (i.e. by eliminating all macroscopic Schrödinger Cat-like superpositions) while preserving (on our corrected pragmatic interpretation) exactly the same experimental predictions as standard quantum mechanics might be regarded as another advantage. But it also of course means that, viewed in its own right, it is only what one might call a *philosophical* resolution in that it leads to a changed ontology (i.e. with no Schrödinger Cat-like macroscopic superpositions) from that of standard quantum mechanics but makes no experimentally distinguishable predictions.

However, it is saved from being ‘merely’ philosophical since the experimental indistinguishability of course only holds for our non-relativistic approximate theory. As is implicit in Section 4 (see also Note <xxvi>) when post-Newtonian corrections to our theory become significant, one would expect experiments along the lines of the Penrose experiment to detect differences. We next turn to consider at more length what the post-Newtonian corrections to our theory might be like.

As our starting point, we return to the Newtonian time-evolution rule summarized in the first paragraphs of this section. When some details are stripped away, this specifies, for a given zero-time density operator $\rho(0)$ with position space representation as in (39) with t set to zero, a map

$$t \mapsto \rho(t) \tag{41}$$

from the full real time-line (i.e. the full set of positive and zero and negative times) to the space of density operators on the relevant many body Hilbert space $\mathcal{H}_{\text{matter}}$.

This rule will arise as the solution to a suitable master equation. We won't write this down explicitly; it is straightforward to obtain it by differentiating Equation (39) with respect to time just as (9) is obtained by differentiating the one-body special case (8). Instead, we will write it schematically as

$$\dot{\rho} = -i[H, \rho] + \textit{Newton} \rho \quad (42)$$

where *Newton* is the appropriate operator which acts on the set of density operators. It will be a 'master equation of MDM type' in the terminology introduced in Note <x>.

In the post-Newtonian regime, we would expect the nature of the dynamics of the physical density operator $\rho(t)$ (i.e. of the trace over $\mathcal{H}_{\text{gravity}}$ of the time-evolving pure total density operator as explained in Subsection 1.1) to be qualitatively different (as we are about to discuss) but we would still expect it to admit of a description in terms of a map of the form of (41) – i.e. from the full real time-line to the same space of many-body density operators. However, unlike (40) and its associated master equation (42), one expects a new sort of rule determining the time-evolution of $\rho(t)$ which incorporates the back-reaction on our many-body matter from radiated gravitons. We would not necessarily expect this new rule to obey an exact master equation but, in suitable models, we would expect there to be a preferred zero-moment of time at which the entropy is a minimum – the many-body post-Newtonian counterpart to the moment (t_{min} in (37)) of minimum entropy in our one-dimensional Gaussian model example of Section 6 – and, while, of course, one now expects the rule which determines $\rho(t)$ at other times to depend on the underlying total time-zero state and not just on the physically relevant density operator $\rho(0)$ (which is its partial trace over gravity – see the beginning of Subsection 1.1), we would expect, in view of the considerations discussed in Note <ix>, that, *for positive times* $\rho(t)$ will *approximately* obey a master equation of schematic form

$$\dot{\rho} = -i[H, \rho] + \textit{Newton} \rho + \textit{postNewton} \rho \quad (43)$$

where *postNewton* is approximately of GKS/L form while, *for negative times* it will approximately obey a similar master equation but with the term *postNewton* replaced by a term *antipostNewton* which is of anti-GKS/L form <ix>. The reason is that (for positive times) the post Newtonian terms will arise by, at each time, tracing the full matter-gravity (pure) density operator over modes of the gravitational field which are *radiative* (see <ix>), and, unlike the non-radiative modes which, in virtue of being 'dragged around' by the matter are responsible for the special features of *Newton* as we discussed in Subsection 4.1, Section 6 and Note <x>, radiative modes will, of course, be expected to fly away from the matter. (And there is of course a similar argument for the expected anti-GKS/L property for negative times.) Moreover, while the terms *postNewton* and *antipostNewton* are expected to be negligibly tiny in comparison to *Newton* in the Newtonian regime (see Note <xxvi>), we would expect that, as one moves away from the Newtonian regime, they will become comparable in size, and as one approaches an extreme relativistic regime, they will be much more important. So one might say that the non-relativistic approximate theory, which is the main subject

of the present paper, is expected to just be the small non-relativistic tip of a large relativistic iceberg! (Of course, the possibility of describing ρ as a many-particle density operator will presumably break down at some point and need to be replaced e.g. by a description in terms of quantum fields or strings etc.)

We next turn to discuss the issue of which systems deserve to be considered *closed* for our various purposes. As we have indicated towards the beginning of Subsection 1.2 and warned in several places above, it was already concluded in [2] that, at least as far as the concept of entropy is concerned, ordinary small laboratory-sized non-relativistic systems cannot be considered closed. We shall next recall the reasoning behind this conclusion and shall then discuss whether or not such a conclusion should also apply to the other matters under discussion here – notably our conclusions about the Penrose (and other) experiment(s) and our tentative interpretation in terms of events.

By a well-known discussion (see e.g. the article by Joos in [26]) based on the argument that there will always inevitably be quantum mechanical entanglement over widely spatially separated subsystems, one can argue that it only really can make exact sense to regard a subsystem of the universe as closed and to regard its underlying state as describable by a pure (i.e. vector) state if the subsystem is the entire universe. Indeed, at a classical level, it was famously pointed out by Emil Borel in 1914 that the motion of a few grams of material on the star, Sirius, would, just because of its gravitational interaction, after a short time, significantly change the configuration of a gas in a box on the Earth. Similar considerations at a quantum level (and for not-necessarily gravitational interactions) make it plausible that, in a quantum description, in the actual state of the universe there will be a considerable degree of entanglement between a bit of gas in a box on the Earth and a bit of gas in the star Sirius etc. (For a discussion of all this, see for example Chapter 3 by E. Joos in [26])

However, one can ask whether it is legitimate to assume a given system is closed for the purposes of some given calculation – if one is content with a good approximate (rather than an exact) answer. We know from experience that, for many questions, such an assumption can be successful for laboratory-sized systems. Otherwise standard quantum mechanical calculations would not be anything like as successful as they actually are. However, if our purpose is the explanation, on the hypotheses of [1], and within the approximate Newtonian framework of [2] of how e.g. a box of gas in the laboratory can have an entropy comparable to the usual entropy one would calculate for a thermal equilibrium state at say room temperature, then it appears, from the preliminary model calculations reported in [2] that this cannot be done: The typical entropies we calculate (see [2]), while non-zero, are orders of magnitude smaller. Our conclusion from this is not that the theory of [2] or the calculations are wrong, but rather that hot boxes of gas in the laboratory cannot be regarded as closed systems. Rather, as we discuss further in Note <xii> (see also Figures 4 and 5) we conclude that the entropy of such systems needs to be explained by regarding them to be open systems and is due to tracing over a total environment, part of which is gravitational, but the main part of which is essentially the same matter environment which (on one standard view about

the origin of the entropy of gases in boxes – see the last paragraph of Note <xii>) would usually be held responsible for their entropy.

It then becomes an important question, whether, to get physically correct values for the entropy of such laboratory-sized systems and to get physical thermodynamic behaviour (i.e. a ‘Second-Law’ result to the effect that the entropy of a closed system always increases) it could suffice to regard as closed, say, the solar system, or whether one needs to include our whole galaxy or whether one actually needs to go as far as the entire universe. We have argued in [2] and will argue further below that any answer to this question must involve enlargement to a system, parts of which are essentially relativistic in nature (the remark in Note <xii> about total closed systems which contain black holes is also relevant to this conclusion). Thus the galaxy or the entire universe would seem to be the more likely answers. And if the galaxy turns out to be a valid answer, then one might speculate that if it is true, as is currently believed likely, that every large galaxy contains a supermassive black hole, then this fact may play a role in this.

What we have studied in [2] and in the present paper are small laboratory-sized model systems on the (we have concluded above as far as entropy is concerned) counterfactual assumption that they are closed. While not expecting their entropies to be physically realistic, we can endeavour to learn lessons about the mechanisms which give rise to entropy and about the way that entropy tends to vary with time. In [2] we considered a number of laboratory-sized model static systems and calculated their entropies on this assumption, and showed, in particular, that these entropies, while (as we mentioned above) small, appear to correlate, to some extent, with their degree of ‘matter clumping’. In Section 6 here, we have augmented these results with some first results on the time-dependence of entropy and found, for our Gaussian model to first order in κ (describing our free bead-on-a-wire model) a two-sided entropy increase result whereby there is a moment of minimum entropy while, as a result of ‘spreading of the would-be wave-packet’, entropy increases for times on either side of this moment (i.e. it decreases before this moment and increases after it). For more ‘realistic’ small non-relativistic many-body systems we would expect there to be a similar two-sided entropy-increase result in suitable circumstances, but we expect the mechanism behind it to involve not only spreading of the would-be wave packet, but also, and in fact often mainly, ‘dynamical clumping’, by which we mean changes in time of the amount of matter clumping. Turning to the post-Newtonian regime, we would still expect to obtain a two-sided entropy-increase result in suitable models, but we expect the way one will understand mathematically the origin of this result will be different (and this will be traceable to the different behaviour of the non-radiative and radiative modes which will be responsible for the two regimes – see <ix> and <x>). In the case of our Gaussian model and (one assumes) in the case of non-relativistic many-body systems, the demonstration of such a two-sided entropy increase result depended/will presumably depend on a study of the relevant single exact master equation of ‘MDM’ type (see after (42) which describes the dynamics of $\rho(t)$). In the post-Newtonian regime,

one expects, instead, that after choosing a suitable model for a low-entropy initial $\rho(0)$, the demonstration of a two-sided entropy increase result will depend on a study of an appropriate approximate master equation of form (43) for positive times and its negative-time counterpart for negative times. To the extent that one can neglect the *Newton* terms in these (cf. our ‘tip of iceberg’ remark in the paragraph containing (43)) the question will be whether or not the GKS/L term *postNewton* satisfies the criterion of Benatti-Narnhofer [36] discussed in Note <ix> (etc. for negative times).

The interpretation of such two-sided entropy increase results will of course be that, while, theoretically, they represent possible dynamics for all times, positive and negative, in actual physical application, only the positive-time behaviour will be relevant and the negative time part should simply be discarded. Either one has in mind that one’s closed system is modelling the whole universe in which case $\rho(0)$ will model the initial state of the universe, or, if one is interested in a model which idealizes some subsystem of the universe as closed (e.g. a star collapsing to a black hole in an otherwise empty universe as is done in Note <v>) then $\rho(0)$ will represent the initial state of that system and its low entropy can either be thought of as an assumption of the model or alternatively, perhaps, as ultimately explainable outside the model as traceable back, via processes in the wider universe, to the low entropy of the initial state of the universe. (Cf. the notion of ‘branch systems’ in [30] although note that, in view of our remarks above about how large closed systems may have to be, it is a delicate question if/when/how this notion is applicable here.)

Of course, the only result we have proven here is our two-sided entropy result, obtained in Section 6, for our one-dimensional Gaussian model, interpretable in terms of a model closed system consisting of our free non-relativistic bead on a wire. This is of course a thoroughly counterfactual model closed system but nevertheless our result is of value as a first entropy-increase result on the hypothesis of [1]. If that hypothesis and the extra assumptions involved in [2] are correct, then that model will, as we have indicated above, be of some significance as one extreme tip (it concerns only one single small non-relativistic bead) of a non-relativistic many-body tip (cf. the paragraph containing (43)) of a full relativistic iceberg which is the Second Law of Thermodynamics!

The other side of the coin to entropy increase (on the hypothesis of [1]) is of course dynamical *decoherence* since, after all, aside from its role in thermodynamics, our S can be thought of as a measure of the ‘amount of decoherence’ of our closed system. So, if our expectations concerning results of Second-Law type for suitable (say, laboratory-sized non-relativistic) models are fulfilled, then we will be able to say that, for suitable ‘time-zero’ initial total states, the amount of decoherence will always increase. Thus, returning to our comparison with theories of ‘collapse model’ type, we will be able to say that our non-relativistic approximate theory shares with ‘collapse models’ (i.e. in addition to eliminating Schrödinger Cat-like superpositions as discussed above) the ability to predict a continual process of decoherence. (The way that GRW does this is mentioned in Note <ix>.) However, the comparison of the two sorts of theories has to be more complicated than that, since, as we have argued above, small laboratory

systems cannot, on our theory be regarded, for the purposes of calculating entropy, as closed. Regarding them, instead, as open (as discussed in Note <xii>) we see that, instead, our theory (i.e. of [2]) entails, as far as its predictions of decoherence are concerned, essentially the same predictions as (see Note <v>) the usual ‘environment induced decoherence paradigm’. (See however the next-but-two paragraph.)

Turning to our tentative notion of ‘events’ defined according to the passage labelled (B) in Section (1.2), we have to expect that, for related reasons to those given above, the detailed structure of the set of events and their probabilities and of the way these change with time will differ considerably, for a large relativistic closed system, from the structure etc. of this set etc. when one analyses simple non-relativistic models of closed systems as in Section 5. Nevertheless, the results of Section 5 are still of interest as a model within which one can attempt to explore what might be the mechanism by which such an events-based interpretation might be able to supplant the stop-gap naive and corrected pragmatic interpretations discussed in Sections 2 and 4. What one hopes for here is to have a theory which explains how quantities such as position and momentum and parity etc. which are treated as ‘observables’ in our naive and corrected pragmatic interpretations *emerge* in some suitable sense from an events interpretation. Our discussion, in Section 5 (see also Note <xxvii>) of the relationship between equations (20 and (36) in the context of our one-dimensional Gaussian model, appears to give a promising indication, for the example of ‘parity’ of how this can happen and the challenge which remains is to study whether an explanation along the lines indicated by that example, of the emergence of less primitive observables in more complicated (many-body) examples.

We emphasize that even if a theory along such lines for how observables emerge from our events interpretation is possible, it is not obvious that it will be possible staying within the context of small non-relativistic model closed systems. After all, we argued above that for questions related to entropy and thermodynamics, such models need to be rejected in favour of models of large relativistic closed systems. However the small success of our discussion of parity encourages us to hope that, for these different purposes, our small non-relativistic models may suffice. Indeed, it seems conceivable that, even though a different, and perhaps richer and ultimately arguably more ‘physical’, theory of events would be had for large relativistic models of closed systems and by treating actual ordinary laboratory-sized systems as open subsystems of these, the theory obtained by modelling actual ordinary (non-relativistic) laboratory-sized systems as closed may, in itself, yield a viable events-based replacement for standard quantum mechanics.

Concomitantly, one might possibly argue that, returning to our comparison between our theory and ‘collapse models’, if we treat a given small non-relativistic laboratory-sized system as if it is ‘closed’ for the purposes of calculating its entropy, S , then, while the resulting S should not be taken seriously as having anything to do with thermodynamics or ‘true extent of decoherence’, it might still be an interesting measure of ‘decoherence relevant to the model’ and might, under certain circumstances, always

increase.

Finally, returning to our naive and pragmatic interpretations which we discussed at the outset, we return to the question whether it is justified, for the results which were obtained within them (including our Position Measurement Theorem and including our analysis of the Penrose experiment) to be taken to be physically relevant. The question is whether such matters can (unlike what we argued above for questions of entropy and thermodynamics) be discussed in the context of the models of small non-relativistic closed systems which we have considered here. Another way of putting this is to ask whether, e.g. in our discussion of the Penrose experiment, we can ignore entanglement of our small model closed systems with the environment. The answer would seem to be that it is justified since (cf. Note <xii>) the issue of entanglement of such systems with the environment within our theory is more or less the same as what would usually be regarded as the issue of the entanglement of such systems with the environment and this is, of course (see especially [9]) precisely the issue which makes the Penrose experiment so difficult (but not impossible) to perform and which the experiment is designed to overcome.

(A brief summary-statement of some of the main conclusions from the above discussion with an emphasis on comparing and contrasting our Newtonian approximate theory with ‘collapse models’ such as GRW [13] was given towards the end of the Introduction.)

8. Acknowledgments

BSK would like to thank The Leverhulme Foundation for a Leverhulme Fellowship (RF&G/9/RFG/2002/0377) from October 2002 to June 2003 during the course of which an important part of this research was done. BSK’s research was also partially supported by PPARC grant ‘Quantum Physics of Fields, Boundaries and Gravitation’ (Sept 2004 to Aug 2006). VA thanks PPARC for a research studentship from October 2002 to September 2005.

9. Appendix: Approximate diagonalisation of the density operator in the one-dimensional Gaussian and some other, related, Models

Given ρ as in (13)

$$\rho(x, x') = \rho_0(x, x')e^{-\kappa(x-x')^2},$$

where

$$\rho_0(x, x') = \psi(x)\psi^*(x'),$$

our aim is to find the eigenvectors and eigenvalues of ρ . I.e. we wish to solve

$$\rho|\phi\rangle = \lambda|\phi\rangle, \tag{44}$$

for $|\phi\rangle$ and λ . This could be studied exactly, but we content ourselves here with solving the simpler problem where ρ is replaced by its first order in κ form, which shall call ρ^κ ,

$$\begin{aligned}\rho^\kappa(x, x') &= \rho_0(x, x') (1 - \kappa(x - x')^2) = \rho_0(x, x') (1 - \kappa x^2 + 2\kappa x x' - \kappa x'^2) \\ &= \rho_0(x, x') + \kappa \rho_1(x, x'),\end{aligned}$$

where

$$\rho_1(x, x') = -\kappa \psi(x) \psi^*(x') (x - x')^2$$

and the resulting eigenvalue equation is solved to first order in κ . It is useful to note, that, in Dirac notation, ρ^κ may be written

$$\rho \simeq \rho_0 + \kappa \rho_1 = |\psi\rangle\langle\psi| + \kappa (-x^2|\psi\rangle\langle\psi| + 2x|\psi\rangle\langle\psi|x - |\psi\rangle\langle\psi|x^2). \quad (45)$$

Taking this as our starting point, we can make the following perturbation-theoretic argument: If κ is small, we expect ρ , and hence also ρ^κ to have an eigenvector close to $|\psi\rangle$, and that its eigenvalue will be close to unity. Thus we write this first eigenvector as

$$|\phi_1\rangle = |\psi\rangle + \kappa|\psi_\perp\rangle, \quad (46)$$

where $|\phi_1\rangle$ is not yet normalized and we assume the orthogonality condition,

$$\langle\psi|\psi_\perp\rangle = 0$$

holds. As $\text{tr } \rho = \text{tr } \rho^\kappa = 1$, and ρ is positive (we will confirm that ρ^κ is positive for sufficiently small κ below) we expect that the eigenvalue of $|\phi_1\rangle$ will be slightly less than unity so we write

$$\rho^\kappa|\phi_1\rangle = (1 - a\kappa)|\phi_1\rangle, \quad (47)$$

where a is to be determined. From (45), (46), and (47) we get

$$\rho^\kappa|\phi_1\rangle = (\rho_0 + \kappa\rho_1)(|\psi\rangle + \kappa|\psi_\perp\rangle) = (1 - a\kappa)(|\psi\rangle + \kappa|\psi_\perp\rangle). \quad (48)$$

We can obtain $|\psi_\perp\rangle$ from (48) by expanding out the brackets to give

$$|\psi_\perp\rangle = (a + \rho_1)|\psi\rangle,$$

and substituting the expression for ρ_1 , given in (45), into the above equation gives

$$|\psi_\perp\rangle = a|\psi\rangle - x^2|\psi\rangle + 2\langle x\rangle x|\psi\rangle - \langle x^2\rangle|\psi\rangle. \quad (49)$$

By only keeping terms up to order κ in (48) we get

$$\kappa\rho_1|\psi\rangle = -a\kappa|\psi\rangle + \kappa|\psi_\perp\rangle \quad (50)$$

Now acting on (50) with $\langle\psi|$ it can be seen that

$$a = -\langle\psi|\rho_1|\psi\rangle = 2(\Delta x)^2, \quad (51)$$

where

$$\Delta x = (\langle\psi|x^2|\psi\rangle - \langle\psi|x|\psi\rangle^2)^{\frac{1}{2}}.$$

Note that the above is the uncertainty in position of the would-be wave function (as we call it in Subsection 1.1). So we are now in a position to write an explicit expression (not normalized) for our first eigenvector of ρ to first order in κ , using (46), (49), and (51)

$$|\phi_1\rangle = (1 + 2\kappa(\Delta x)^2) |\psi\rangle - \kappa x^2 |\psi\rangle + 2\kappa\langle x\rangle x |\psi\rangle - \kappa\langle x^2\rangle |\psi\rangle, \quad (52)$$

and its eigenvalue, from (47) and (51), is

$$\lambda_1 = 1 - 2\kappa(\Delta x)^2. \quad (53)$$

Until now we have made no assumptions about the state (i.e. would-be wave function) $|\psi\rangle$, but in order to have a simple form for the remaining eigenvectors we shall specialize from now on to the cases where ψ is either an even function of x (i.e. $\psi(x) = \psi(-x)$) or an odd function of x (i.e. $\psi(x) = -\psi(-x)$). Using this assumption in (52) and (53) we have that

$$|\phi_1\rangle = (1 + \kappa\langle x^2\rangle) |\psi\rangle - \kappa x^2 |\psi\rangle,$$

and

$$\lambda_1 = 1 - 2\kappa\langle x^2\rangle.$$

Next we note that any eigenvector $|\phi\rangle$ must (to order κ) satisfy

$$\begin{aligned} \rho^\kappa |\phi\rangle &= |\psi\rangle\langle\psi|\phi\rangle - \kappa x^2 |\psi\rangle\langle\psi|\phi\rangle \\ &\quad + 2\kappa x |\psi\rangle\langle\psi|x|\phi\rangle - \kappa |\psi\rangle\langle\psi|x^2|\phi\rangle = \lambda |\phi\rangle. \end{aligned} \quad (54)$$

from which it can be seen that any non-zero eigenvector of ρ^κ , must be a linear combination of $|\psi\rangle$, $x|\psi\rangle$ and $x^2|\psi\rangle$. Using this observation as a clue, it is easy to see that there is another eigenvector, $|\phi_2\rangle$, with a non-zero eigenvalue, and that it has the simple form

$$|\phi_2\rangle = x|\psi\rangle, \quad (55)$$

with an eigenvalue equal to $2\kappa\langle x^2\rangle$, which can be seen from (55) and (45).

Since the sum of the eigenvalues we have found so far is 1 ($= \text{tr } \rho = \text{tr } \rho^\kappa$) and on the assumption that ρ^κ , like ρ is positive, it must be that the orthogonal complement to the (two-dimensional) subspace spanned by the eigenvectors we have found so far must consist entirely of eigenvectors with eigenvalue zero (to first order in κ). Thus, to first order in κ , we have found all eigenvectors and eigenvalues of ρ^κ and (to this order) these are obviously the same as for ρ .

We remark that an alternative derivation of the same result, and confirmation that, for sufficiently small κ , ρ^κ is positive, may be had by noting that (still with our specialization to ψ either even or odd) to first order in κ , (45) may be written

$$\rho = |(1 - \kappa x^2 \psi)\rangle\langle(1 - \kappa x^2 \psi)| + 2\kappa |(x\psi)\rangle\langle(x\psi)|$$

This is in the canonical form,

$$\rho = \sum_n b_n |\psi_n\rangle\langle\psi_n|$$

except that the vectors which appear, while orthogonal, are not normalized. To remedy this, we easily have (to first order):

$$|(1 - \kappa x^2 \psi)| = 1 - \kappa \langle \psi | x^2 \psi \rangle = 1 - a\kappa$$

(where, for convenience, we resume our abbreviation of $\langle \psi | x^2 \psi \rangle$ by a) so that (again, all equalities are to first order) the normalized vector is

$$(1 + a\kappa)(1 - \kappa x^2 \psi) = 1 + a\kappa - \kappa x^2 \psi$$

and also

$$|x\psi| = a^{1/2}.$$

We can now replace the vectors by their normalized forms at the expense of introducing prefactors equal to the square of their norms. Thus

$$\rho = (1 - 2a\kappa)|(1 + a\kappa - \kappa x^2 \psi)\rangle\langle(1 + a\kappa - \kappa x^2 \psi)| + 2a\kappa|(a^{-\frac{1}{2}}x\psi)\rangle\langle(a^{-\frac{1}{2}}x\psi)|,$$

from which we can read off the eigenvectors and eigenvalues. We see that these are the same as those we obtained (by a rather different route) above.

9.1. Diagonalizing a tensor product state in one dimension

Consider the tensor product state

$$|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \dots \otimes |\psi_N\rangle,$$

and the density matrix

$$\begin{aligned} \rho(x_1, \dots, x_N; x'_1, \dots, x'_N) &= \Psi(x_1, \dots, x_N) \Psi^*(x'_1, \dots, x'_N) e^{-\kappa(x_1 + x_2 + \dots + x_N - x'_1 - x'_2 - \dots - x'_N)^2} \\ &= \Psi(x_1, \dots, x_N) \Psi^*(x'_1, \dots, x'_N) \left(1 - \kappa [(x_1 + x_2 + \dots + x_N) - (x'_1 + x'_2 + \dots + x'_N)]^2 \right), \end{aligned}$$

to $O(\kappa)$. (From now on, we shall not always bother to mention that we are working to order κ and we shall not bother to distinguish between ρ and the counterpart to what we called above ρ^κ .) In Dirac notation this is

$$\begin{aligned} \rho &= |\Psi\rangle\langle\Psi| - \kappa(x_1 + \dots + x_N)^2 |\Psi\rangle\langle\Psi| \\ &\quad + 2\kappa(x_1 + \dots + x_N) |\Psi\rangle\langle\Psi| (x_1 + \dots + x_N) - \kappa |\Psi\rangle\langle\Psi| (x_1 + \dots + x_N)^2. \end{aligned} \quad (56)$$

From the previous example we are tempted to hope that there will once again be only be two eigenstates of our density matrix, and that they will have a similar form. We know that there is only one eigenstate of the density matrix $\rho_0 = |\Psi\rangle\langle\Psi|$, namely $|\Psi\rangle$. Thus we expect that there will be an eigenstate close to $|\Psi\rangle$ with eigenvalue slightly less than unity. Let us write our first eigenvector as

$$|\Phi_1\rangle = |\Psi\rangle + \kappa |\Psi_\perp\rangle \quad (57)$$

where $\langle\Psi|\Psi_\perp\rangle = 0$. Then

$$\begin{aligned} \rho|\Phi_1\rangle &= [|\Psi\rangle\langle\Psi| - \kappa(x_1 + \dots + x_N)^2 |\Psi\rangle\langle\Psi| + 2\kappa(x_1 + \dots + x_N) |\Psi\rangle\langle\Psi| (x_1 + \dots + x_N) \\ &\quad - \kappa |\Psi\rangle\langle\Psi| (x_1 + \dots + x_N)^2] (|\Psi\rangle + \kappa |\Psi_\perp\rangle) = (1 - \kappa a) |\Phi_1\rangle \\ &= (1 - \kappa a) (|\Psi\rangle + \kappa |\Psi_\perp\rangle). \end{aligned} \quad (58)$$

Equating the $O(\kappa)$ terms in the above equation yields

$$\begin{aligned} - (x_1 + \dots + x_N)^2 |\Psi\rangle + 2(x_1 + \dots + x_N) |\Psi\rangle \langle \Psi | (x_1 + \dots + x_N) |\Psi\rangle \\ - |\Psi\rangle \langle \Psi | (x_1 + \dots + x_N)^2 |\Psi\rangle = -a |\Psi\rangle + |\Psi_\perp\rangle. \end{aligned} \quad (59)$$

In order to proceed and to be able to find the remaining eigenvector we assume that the wave functions $\psi_j(x)$ are either even or all odd for all $j \in 1 \dots N$, bearing in mind that $|\Psi\rangle = |\psi_1\rangle \otimes \dots \otimes |\psi_N\rangle$. That is to say that for each j we have either $\psi_j(x) = \psi_j(-x)$ or $\psi_j(x) = -\psi_j(-x)$. Assuming this and acting on both sides of the above equation with $\langle \Psi |$ gives

$$a = 2 (\langle x_1^2 \rangle + \dots + \langle x_N^2 \rangle). \quad (60)$$

Substituting (60) into (58) we are able to deduce $|\Psi_\perp\rangle$, and hence the first eigenstate (by substituting $|\Psi_\perp\rangle$ into (57)),

$$|\phi_1\rangle = [1 + \kappa (\langle x_1^2 \rangle + \langle x_2^2 \rangle + \dots + \langle x_N^2 \rangle)] |\Psi\rangle - \kappa (x_1 + x_2 + \dots + x_N)^2 |\Psi\rangle,$$

along with its eigenvalue

$$\lambda_1 = 1 - 2\kappa (\langle x_1^2 \rangle + \langle x_2^2 \rangle + \dots + \langle x_N^2 \rangle).$$

It can be shown that there is only one other eigenvector,

$$|\phi_2\rangle = (x_1 + x_2 + \dots + x_N) |\Psi\rangle,$$

with eigenvalue

$$\lambda_2 = 2\kappa (\langle x_1^2 \rangle + \langle x_2^2 \rangle + \dots + \langle x_N^2 \rangle),$$

which is similar to the previous example.

9.2. Diagonalizing the density operator in three dimensions

Let the wave function $\psi(\mathbf{x})$ satisfy the following symmetry conditions

$$\psi(x, y, z) = \psi(-x, y, z) = \psi(-x, -y, z) = \psi(-x, -y, -z) = \quad (61)$$

$$\psi(-x, y, -z) = \psi(x, -y, z) = \psi(x, -y, -z) = \psi(x, y, -z), \quad (62)$$

and be such that it can be written as a product of three independent functions,

$$\psi(x, y, z) = \alpha(x)\beta(y)\gamma(z).$$

Consider the density operator

$$\rho(\mathbf{x}, \mathbf{x}') = \rho_0(\mathbf{x}, \mathbf{x}') e^{-\kappa(\mathbf{x} - \mathbf{x}')^2} \simeq \rho_0(\mathbf{x}, \mathbf{x}') - \kappa(\mathbf{x} - \mathbf{x}')^2 \rho_0(\mathbf{x}, \mathbf{x}')$$

where $\rho_0(\mathbf{x}, \mathbf{x}') = \psi(\mathbf{x})\psi^*(\mathbf{x}')$. We wish to find the eigenvectors and eigenvalues of ρ to $O(\kappa)$, which means that we want to find $\phi(\mathbf{x})$ and λ that satisfy

$$\int d^3 \mathbf{x}' \rho(\mathbf{x}, \mathbf{x}') \phi(\mathbf{x}') = \lambda \phi(\mathbf{x}). \quad (63)$$

In Dirac notation our density operator, to first order in κ , is

$$\rho = |\psi\rangle \langle \psi| - \kappa \mathbf{x}^2 |\psi\rangle \langle \psi| + 2\kappa \mathbf{x} |\psi\rangle \langle \psi| - 2\kappa |\psi\rangle \langle \psi| \mathbf{x}^2.$$

To find the first eigenvector we use a similar perturbative argument to the previous one to deduce that there must be an eigenvector close to $|\psi\rangle$ with eigenvalue slightly less than unity, i.e.

$$\phi_1(\mathbf{x}) = \psi(\mathbf{x}) + \kappa\psi_\perp(\mathbf{x}),$$

or in Dirac notation

$$|\phi_1\rangle = |\psi\rangle + \kappa|\psi_\perp\rangle. \quad (64)$$

We want to find $|\psi_\perp\rangle$ such that $|\phi_1\rangle$ is an eigenvector of ρ . To do this we first write

$$\begin{aligned} \rho|\phi_1\rangle &= (|\psi\rangle\langle\psi| - \kappa\mathbf{x}^2|\psi\rangle\langle\psi| + 2\kappa\mathbf{x}|\psi\rangle\langle\psi|\mathbf{x} - \kappa|\psi\rangle\langle\psi|\mathbf{x}^2) (|\psi\rangle + \kappa|\psi_\perp\rangle) \\ &= (1 - a\kappa)(|\psi\rangle + \kappa|\psi_\perp\rangle). \end{aligned} \quad (65)$$

From this, and using the symmetry restrictions we imposed on $|\psi\rangle$, we obtain $|\psi_\perp\rangle$,

$$|\psi_\perp\rangle = -(\mathbf{x}^2 + \langle\mathbf{x}^2\rangle)|\psi\rangle. \quad (66)$$

By acting on both sides of (65) with $\langle\psi|$ we find that $a = 2\langle\mathbf{x}^2\rangle$. Thus our first eigenvalue is

$$\lambda_1 = 1 - 2\langle\mathbf{x}^2\rangle$$

and the corresponding eigenvector (obtained by substituting (66) into (64)),

$$|\phi_1\rangle = (1 - \kappa\langle\mathbf{x}^2\rangle)|\psi\rangle - \kappa\mathbf{x}^2|\psi\rangle.$$

We claim that the following are eigenvectors of ρ ,

$$\phi_2(\mathbf{x}) = x\psi(\mathbf{x}); \quad \phi_3(\mathbf{x}) = y\psi(\mathbf{x}); \quad \phi_4(\mathbf{x}) = z\psi(\mathbf{x}),$$

with respective eigenvalues

$$\lambda_2 = 2\kappa\langle x^2\rangle; \quad \lambda_3 = 2\kappa\langle y^2\rangle; \quad \lambda_4 = 2\kappa\langle z^2\rangle.$$

To show that $\phi_2(\mathbf{x})$ is an eigenvector we simply show that it satisfies (63) through

$$\begin{aligned} \int d^3\mathbf{x}' \rho(\mathbf{x}, \mathbf{x}') \phi_2(\mathbf{x}') &= \int dx' dy' dz' \alpha(x) \beta(y) \gamma(z) \alpha^*(x') \beta^*(y') \gamma^*(z') x' \alpha^*(x') \beta^*(y') \gamma^*(z') \\ &\quad - \kappa \int dx' dy' dz' (x^2 + y^2 + z^2 - 2xx' - 2yy' - 2zz' + x'^2 + y'^2 + z'^2) \\ &\quad \alpha(x) \beta(y) \gamma(z) \alpha^*(x') \beta^*(y') \gamma^*(z') x' \alpha^*(x') \beta^*(y') \gamma^*(z'). \end{aligned}$$

The symmetry restrictions on $\psi(x, y, z)$ imply that the first term in the above equation vanishes. Noting this and then expanding out the second term gives

$$\begin{aligned} \int d^3\mathbf{x}' \rho(\mathbf{x}, \mathbf{x}') \phi_2(\mathbf{x}') &= -\kappa\psi(x, y, z)(x^2 + y^2 + z^2) \\ &\quad \times \int dx' dy' dz' \alpha^*(x') \beta^*(y') \gamma^*(z') x' \alpha(x') \beta(y') \gamma(z') - \kappa\psi(x, y, z) \\ &\quad \times \int dx' dy' dz' \alpha^*(x') \beta^*(y') \gamma^*(z') (x'^2 + y'^2 + z'^2) x' \alpha(x') \beta(y') \gamma(z') \\ &\quad + 2\kappa\psi(x, y, z) \int dx' dy' dz' \alpha^*(x') \beta^*(y') \gamma^*(z') (xx' + yy' + zz') x' \alpha(x') \beta(y') \gamma(z'). \end{aligned}$$

Once again the first two terms of the above equation vanish and we are left with

$$\int d^3\mathbf{x}' \rho(\mathbf{x}, \mathbf{x}') \phi_2(\mathbf{x}') = 2\kappa \langle x^2 \rangle x \psi(x, y, z) = 2\kappa \langle x^2 \rangle \phi_2(\mathbf{x}),$$

which shows that $\phi_2(\mathbf{x})$ is an eigenvector. It is easy to show that $\phi_3(\mathbf{x})$ and $\phi_4(\mathbf{x})$ are eigenvectors by a similar argument. This example, with its artificial symmetry restrictions on the wave function $\psi(\mathbf{x})$, serves as an indication that there are more eigenstates of ρ when investigating higher-dimensional systems.

10. Notes

- (i) The problems and puzzles which concern quantum physics in general include:

measurement Those usually collected under the heading ‘the measurement problem in quantum mechanics’.

entropy The puzzling contradiction between the traditional understanding of the entropy of a general closed system in terms of (subjective) ‘coarse-graining’ and our understanding, since the work of Bekenstein [33] and Hawking [34] in the 1970s that the entropy of the specific closed system consisting of a black hole (either sitting in and radiating into, an asymptotically flat empty universe or in equilibrium with radiation in a suitable box) is something objective – namely something which is approximately equal to ‘one quarter of the area of the event horizon’.

second law The problem of supplying a microscopic explanation for the second law of thermodynamics, i.e. a microscopic explanation for why the entropy of a closed system always increases.

The problems and puzzles which relate specifically to black holes include:

information loss The ‘information loss puzzle’, which we shall take here to mean the puzzle as to how it can be that, on the one hand, during the process of stellar collapse and black-hole formation and then black-hole evaporation, entropy presumably actually continually increases while, on the other hand, the underlying quantum dynamics of the process is presumably unitary. We remark that what makes this a puzzle is the traditional (but incorrect according to [1] and the present paper) assumption that (in this context) the physical entropy of a quantum mechanical system should be identified with the von Neumann entropy of its (total) density operator – the puzzle then arising because this quantity, being a unitary invariant, must therefore remain constant under a unitary time-evolution and, in fact, if the (initial) state is a vector state, be zero (at all times).

coincidence The problem (raised and solved in [1]) of explaining the coincidence that the thermodynamic entropy $S_{\text{thermodynamic}}(M)$ of a (assumed for simplicity spherically symmetric, neutral) black hole of mass M turns out to

have (approximately) the same value – namely $1/4\pi M^2$, i.e. ‘one quarter of the area of the event horizon’ – as the gravitational entropy $S_{\text{Gibbons-Hawking}}^{\text{matterless}}(T)$ of a Gibbs state of matterless gravity at temperature T when T is taken to be the Hawking temperature $T = 1/8\pi M$.

Here, what we mean by the thermodynamic entropy $S_{\text{thermodynamic}}(M)$ is the value of S implied by the standard thermodynamic relation

$$dE = TdS$$

when we identify E with M and T with $1/8\pi M$ and take S to be 0 when $M = 0$. This is easily calculated to be $4\pi M^2$. On the other hand, what we mean by the gravitational entropy $S_{\text{Gibbons-Hawking}}^{\text{matterless}}(T)$ is the value of S calculated using the standard equilibrium-statistical-mechanics formula

$$S = \ln Z - \beta \partial \ln Z / \partial \beta$$

from the Gibbons-Hawking [31] Euclidean-quantum-gravity partition function $Z(\beta)$, at temperature $T = 1/\beta$, of a Gibbs state of a *matterless* <iii> quantum gravitational field in a suitable (spherical) box. In fact, when (cf. [31]) this partition function is approximated by its zero-loop value $\exp(-I_{\text{classical}})$ with $I_{\text{classical}}$ the classical action, $4\pi M^2$, of the Euclideanized Schwarzschild metric of mass $M = 1/8\pi T$ ($= \beta/8\pi$), the value of S obtained in this way is again easily seen to be $4\pi M^2$. (We note that the matterless gravitational field here is a mathematical construct and, from a physical point of view of course, counterfactual. What we call the Gibbons-Hawking partition function for this is the quantity one would obtain if one were to remove the terms representing matter in the equations in [31] – see the next Note <iii>)

Other puzzles for which our hypothesis offers natural resolutions include the **thermal atmosphere puzzle** and a puzzle which we raise here which we call the **neutron-star entropy puzzle**. We shall say what these are and give our proposed resolutions of them at the end of Note <v>.

- (ii) In an exact unified theory at the Planck energy and higher, one would, of course, not expect there to be a clear demarcation between ‘matter’ and ‘gravity’. But at energies well below the Planck energy, one expects that there will. By ‘matter’ here, we of course mean everything other than gravity. We remark that our hypothesis entails that entropy is a quantity which *emerges* at low energies and will cease to have any meaning at energies around and above the Planck energy.
- (iii) (This Note is really a footnote to Note <i>.) To give more details about the connection with the work of Gibbons-Hawking, and to explain what the traditional assumptions have to say about our coincidence puzzle, we emphasize that our statement of our **coincidence** puzzle refers to the partition function for a Gibbs state of a matterless gravitational field (i.e. ‘pure gravity’) whereas in [31] – see also the review and further discussion in [32] – the partition function actually considered is that for a Gibbs state of gravity together with matter fields. It turns

out that the zero-loop approximation to this is identical to that for matterless gravity discussed in the main text. However the one-and-higher loop corrections will involve the matter as well as the gravity. In these (and many other, later) references the traditional assumption is adopted that it is the Gibbs state of gravity together with appropriate matter which represents the physical total state of a black hole in equilibrium in a box and therefore they find the total physical entropy to be the sum of the same (matterless-gravity) zero-loop contribution (equal to one quarter of the area of the event horizon) discussed here in the main text with a term arising from the one-and-higher-loop matter and gravity part of their partition function which is supposed in those references to represent a correction to the entropy due to the *thermal atmosphere* of the black hole. It has become clear since the early work of Gibbons and Hawking that, due to a quantum-field-theoretic divergence which is, in its turn, due to the infinite coordinate-distance of the horizon from any point outside the horizon when the appropriate (i.e. ‘tortoise’) coordinate is used, this one-loop term is actually impossible to calculate within the Euclidean quantum gravity framework without imposing an ad hoc cut-off near the horizon and its value is therefore in doubt (and, as far as we are aware, this doubt has still not been convincingly removed by any other approach to quantum gravity). For the traditional assumptions to resolve the coincidence puzzle, one would clearly have to assume, as is in fact implied in the Gibbons-Hawking references, that the one-loop ‘thermal atmosphere’ entropy is only a small correction to the zero-loop term, so that the total entropy of the traditionally-described black hole in thermal equilibrium in a box will be well approximated by what we called in Note <i> $S_{\text{Gibbons-Hawking}}^{\text{matterless}}(T)$. This would then formally resolve the **coincidence** puzzle within the traditional assumptions, since, as in the resolution on our hypothesis, both the quantities $S_{\text{Gibbons-Hawking}}^{\text{matterless}}(T)$ (for $T = 1/8\pi M$) and $S_{\text{thermodynamic}}(M)$ would again be understood to approximate the physical entropy of a physical black hole of mass M – in one case in equilibrium in a box and in the other case radiating into empty space. But this is surely a very unsatisfactory and unconvincing resolution of **coincidence** since on the one hand, the quantity $S_{\text{thermodynamic}}(M)$ derives from the thermodynamics of the thermal atmosphere (which, after all is what would escape as Hawking radiation were the walls of the box to be removed) while, on the other hand, the entropy of the thermal atmosphere has to be asserted to be approximately zero in order to make the resolution work! This is of course related to the **thermal atmosphere puzzle** to which, as we explain in Note <v>, our hypothesis also offers a natural resolution. Indeed, on our hypothesis, the entropy of the matter part of the thermal atmosphere is predicted to be approximately equal to $S_{\text{Gibbons-Hawking}}^{\text{matterless}}(T)$ which, up to one-and-higher loop graviton corrections, is $4\pi M^2$. The one-and-higher loop graviton corrections themselves, on the other hand, are plausibly comparable in magnitude to the entropy of just one of the matter fields making up the thermal atmosphere, and therefore might be expected to have a value of the same order of

magnitude as $4\pi M^2/N$ – where N is some appropriate value for the ‘number of matter fields in nature’ – and hence plausibly small.

- (iv) In general, given a pure (i.e. vector) state with density operator $\rho = |\Psi\rangle\langle\Psi|$ on a total Hilbert space, \mathcal{H} , which arises as a tensor product, $\mathcal{H}_m \otimes \mathcal{H}_g$ of two other Hilbert spaces (‘m’ and ‘g’ here could stand for ‘matter’ and ‘gravity’ or could be given a more general interpretation) then one can define the two partial density operators ρ_m on \mathcal{H}_m (the standard partial trace of ρ over \mathcal{H}_g) and ρ_g on \mathcal{H}_g (the standard partial trace of ρ over \mathcal{H}_m) and it is well-known and easy to show that the two von Neumann entropies $S_m = -\text{tr}(\rho_m \ln \rho_m)$ and $S_g = -\text{tr}(\rho_g \ln \rho_g)$ are equal. In modern terminology, an alternative name for these latter (equal) quantities is the *m-g entanglement entropy* of the state ρ .
- (v) Our hypothesis may be understood as a specific variant of the environment-induced decoherence paradigm [25, 26] namely a variant in which the total state of a closed system is taken to be a pure state evolving in time according to a unitary time-evolution. But it has three crucial new features, (1) that the gravitational field has a privileged status as a permanent piece (see Note <xii>) of the total environment, (2) that the entropy which results by tracing over it is regarded as an attribute of the total closed matter-gravity system rather than only of the matter, and (3) that this entropy is identified with the closed system’s physical entropy. With these three new features, one obtains, as far as we are aware, for the first time, a definite notion of decoherence even for closed systems and an objective definition (the matter-gravity entanglement entropy) for the physical entropy of a general closed system, thus offering a resolution to the **entropy** puzzle (see Note <i>). Note of course that this resolution entails that one rejects both the option of identifying the physical entropy of a closed system with the von Neumann entropy of its total density operator ρ_{total} (which would of course be zero if, as our hypothesis assumes, ρ is pure, i.e. a projector onto a single vector) and that one also rejects the traditional attempt to define it in terms of some sort of (subjective) coarse-graining of ρ_{total} .) (For this to be a resolution of the **entropy** puzzle, we of course need to argue that in the specific case of a closed system consisting of a black hole, the matter-gravity entanglement entropy will have the correct value – i.e. (approximately) one-quarter of the area of the event horizon. We postpone our argument for this to the next paragraph.) Moreover, by adopting our new general definition of entropy, one obtains a plausible mechanism for an entropy-increase result for closed systems thus offering resolutions to the puzzles **second law** and **information loss** as we explained in the main text.

Turning to the application of our general hypothesis to the subject of black hole thermodynamics, our hypothesis goes naturally together with a radically different description of quantum black holes from the traditional description. In this new description, the total state of a (say spherically symmetric, neutral) quantum black hole of mass M in equilibrium with radiation in a suitable box is understood to

be a pure state of a quantum-gravitational *closed* system and not, as would be traditionally assumed, a Gibbs state at the Hawking temperature. However, this pure total state is expected to possess, in the relevant low energy description, just the right sort of matter-gravity entanglement for its partial trace ρ_{matter} over $\mathcal{H}_{\text{gravity}}$ to resemble a Gibbs state of the matter fields at the Hawking temperature $T = 1/8\pi M$ in the presence of a background black hole of mass M , and also for its partial trace, ρ_{gravity} , over $\mathcal{H}_{\text{matter}}$, to resemble a Gibbs state of (matterless) gravity at the same Hawking temperature. In fact, it is reasonable to assume, the latter state may be identified with the Gibbs state of a matterless quantum gravitational field – whose partition function may be calculated by the Gibbons-Hawking Euclidean path-integral method summarized in Note <i> by removing the matter terms in [31]. (See Note <iii>.) We remark that of course this extension of our hypothesis entails that the appropriate ‘theory of everything’ will turn out to have states which, at suitably low energies admit of such a description with appropriate matter multiplets etc.

We are now in a position to complete our resolution of the **entropy** puzzle, for, by the result mentioned in Note <iv>, the matter-gravity entanglement entropy of our black hole state must equal the von Neumann entropy of ρ_{gravity} and therefore, in view of the second assumed resemblance mentioned above, must be close to ‘one quarter of the area of the event horizon’ ($4\pi M^2$). Moreover we see that, as a byproduct of this discussion, our hypothesis has led to a natural resolution to the **coincidence** puzzle since, having identified the matter-gravity entanglement entropy with the physical entropy of our black hole, we would also expect it to be the correct ‘ S ’ in the thermodynamic relation ‘ $dE = TdS$ ’.

Our hypothesis also offers a natural resolution to some other puzzling features of the traditional understanding of black hole entropy. Thus, previously, it was unclear to what extent the entropy of a black hole (say in equilibrium with radiation in a box) resides in the quantum geometry of the hole itself, i.e. in the quantum gravitational field, and to what extent it resides in the matter-fields, i.e. in the thermal atmosphere. This is essentially the **thermal atmosphere puzzle** discussed e.g. in [5]. It is particularly puzzling because, on the traditional understanding, it was possible to make arguments (cf. Note <iii>) to the effect that the entropy all resides in the gravitational field but one could also (cf. also e.g. the recent review by Don Page [35]) make counter-arguments to the effect that it all resides in the thermal atmosphere. As a further byproduct to the discussion above, we see that our hypothesis also offers a very neat solution to this puzzle. For if we identify ‘the entropy of the gravitational field’ with the von Neumann entropy of ρ_{gravity} and ‘the entropy of the (matter part of) the thermal atmosphere’ with ρ_{matter} and the entropy of the total state with the matter-gravity entanglement entropy, then, by our resolution to **entropy** (combined with the result mentioned in <iv>) we see that our hypothesis entails that all these three quantities are actually identical. In particular it turns out on our hypothesis that both statements ‘the

entropy resides in the gravitational field’ and ‘the entropy resides in the thermal atmosphere’ (which traditionally were regarded as mutually exclusive possibilities) become simultaneously correct statements!

Another puzzle in the traditional understanding (which we point out here) was that while a black hole (formed say from neutron star collapse) is traditionally understood to have a very large non-zero entropy, the entropy of a neighbouring (not necessarily stationary) state of quantum gravity which consists just of a neutron star (on the verge of collapsing) would, traditionally, be assigned a zero entropy. We call this the **neutron star entropy puzzle**. With our definition (matter-gravity entanglement entropy) any state of any closed system would be expected to have a non-zero entropy and the entropy of such a neighbouring neutron-star state could, quite plausibly, have a value which neighbours (but is a bit less than) the entropy of the black hole it is on the verge of collapsing to. (For preliminary estimates of the entropies, on our hypothesis, of certain simple model matter-gravity systems in the Newtonian limit, see [2]. These have very small values, but, as discussed in Section 7, the Newtonian results are expected only to be “the small non-relativistic tip of a large relativistic iceberg”!)

What our hypothesis tells us about the puzzle **measurement** is of course the subject matter of the rest of this paper.

Finally, we remark that one might regard the fact that our hypothesis allows a more satisfactory definition for the entropy of an open system than traditional ideas do as further evidence for the correctness of our hypothesis. See Note <xii>.

- (vi) As explained in [2], $e^{-D(\mathbf{a}_1, \mathbf{a}_2)}$ is the value of the inner product $\langle g_1 | g_2 \rangle$, in a suitable Hilbert space, $\mathcal{H}_{\text{gravity}}$, for the linearized quantized gravitational field, between two (non-radiative) state vectors, $g_1, g_2 \in \mathcal{H}_{\text{gravity}}$ which are the quantum representations of the static Newtonian gravitational field due to our ball at rest with its centre of mass located at $\mathbf{a}_1, \mathbf{a}_2$ respectively. (As discussed in [2], $\langle g_1 | g_2 \rangle$ turns out to be real.) To understand how the formula (3) arises, it is helpful to first consider a would-be wave function, ψ , which consists of two sharp peaks localized around, say \mathbf{a}_1 and \mathbf{a}_2 ; we shall write

$$\psi = c_1 \psi_1 + c_2 \psi_2 \tag{67}$$

where $c_1, c_2 \in \mathbb{C}$, ψ_1 is a normalized wave function consisting of single sharp peak centred on \mathbf{a}_1 and ψ_2 is a normalized wave function consisting of a single sharp peak centred on \mathbf{a}_2 .

In this case, it is envisaged that a full quantum gravitational description is given by a vector in $\mathcal{H}_{\text{matter}} \otimes \mathcal{H}_{\text{gravity}}$ which takes the form

$$\Psi = c_1 \psi_1 \otimes g_1 + c_2 \psi_2 \otimes g_2.$$

ρ , obtained by tracing $|\Psi\rangle\langle\Psi|$ over $\mathcal{H}_{\text{gravity}}$, is then clearly given by

$$\rho = |c_1|^2 |\psi_1\rangle\langle\psi_1| + c_1 c_2^* \langle g_2 | g_1 \rangle |\psi_1\rangle\langle\psi_2| + c_1^* c_2 \langle g_1 | g_2 \rangle |\psi_2\rangle\langle\psi_1| + |c_2|^2 |\psi_2\rangle\langle\psi_2| \tag{68}$$

and a little thought should suffice to see that, (a) the sharper the two peaks in the would-be wave function, the closer this will be, in its position space representation to (3), (b) for a general would-be wavefunction, one may argue that the formula (3) must still hold by regarding it as a limiting case of a wave function with multiple peaks.

- (vii) The generalization of the Gaussian asymptotic regime to N balls would be relevant e.g. to a superposition of two configurations, each involving N balls, such that, while the balls within each configuration are possibly well separated, the configurations themselves differ only slightly so that each ball in the second configuration is displaced by much less than its radius relative to its position in the first configuration. We remark that the formula easily generalizes to the case of N balls of equal radius R but possibly differing masses in which case D takes the form

$$9(M_{\text{total}}^2/R^2)(x_{\text{cm}} - x'_{\text{cm}})^2$$

where M_{total} is sum of the N masses and x_{cm} and x'_{cm} the centres of mass of the unprimed and primed configurations which constitute the arguments of D . We also remark that the generalization to the many-ball case of the logarithmic asymptotic regime has

$$\exp(-D(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{x}'_1, \dots, \mathbf{x}'_N)) = \prod_{I=1}^N \prod_{J=1}^N \left(\frac{|\mathbf{x}'_I - \mathbf{x}_J| |\mathbf{x}_I - \mathbf{x}'_J|}{|\mathbf{x}_I - \mathbf{x}_J| |\mathbf{x}'_I - \mathbf{x}'_J|} \right)^{-12M_I M_J} \quad (69)$$

where it is to be understood that, in the cases $I = J$, the terms in the denominator $|\mathbf{x}_I - \mathbf{x}_J| |\mathbf{x}'_I - \mathbf{x}'_J|$ are to be replaced by R_I^2 (This is an easily obtained generalization of the formula given ('for simplicity') in [2] in the case where all the M_I and all the R_I are the same. The robustness of the latter formula under changes in shape and graininess of the bodies is discussed in [3].

- (viii) The rule (8) is an automatic consequence (see Note <vi>) of taking the total state for a would-be wave function $\psi(\mathbf{x})$, say with a single sharp peak at $\mathbf{x} = \mathbf{a}$, to be $\psi \otimes g$ where g is the (non-radiative) quantum counterpart to the static Newtonian gravitational field of our ball when its centre of mass is located at \mathbf{a} – irrespective of whether the would-be wave function itself is static or not. Obviously, it amounts to neglecting the radiative part of the gravitational field. This approximation and its generalizations to the many-ball case etc. are however expected to be excellent as long as the classical gravitational radiation from the relevant systems would be negligible. See Note <xxvi> for a quantitative estimate on how good this approximation is likely to be.
- (ix) Ignoring some technicalities, the GKS/L form (often known as 'Lindblad form') for a master equation holds if and only if the master equation integrates up, for positive times, to a 'semi-group of completely positive maps' (sometimes known as a 'quantum dynamical semigroup') on the relevant space of density operators – see [6], [7]. ([6] treated such semigroups of matrices and [7] of bounded

operators.) A prototype (see Section 6) is the BLP master equation (38) [27] $\dot{\rho} = -i[H, \rho] - c[x, [x, \rho]]$, $c > 0$, for a one-particle density operator.

If one has a total Hilbert space, \mathcal{H} , which arises as a tensor product of form $\mathcal{H}_m \otimes \mathcal{H}_g$, and if the dynamics of a total state vector $\Psi(t) \in \mathcal{H}$ evolves according to a (unitary) Schrödinger time-evolution for some Hamiltonian, H , then, if the time-zero state vector $\Psi(0)$ takes the product form $\phi_m \otimes \psi_g$, then one knows ([6]) that for any (positive or negative) time t , the partial trace, $\rho_m(t)$, of $\rho(t) = |\Psi(t)\rangle\langle\Psi(t)|$ over \mathcal{H}_g will arise by the action on the (pure) time-zero density operator $\rho_m(0) = |\phi_m\rangle\langle\phi_m|$ of a ‘completely positive map’ (which will of course depend on ψ_g). (See e.g. also [6] for the definition of ‘completely positive’.)

Moreover, for positive times and if \mathcal{H}_m represents now, say, ‘charged stuff’ and \mathcal{H}_g now, say, ‘radiation’, and if H is such that the ‘produced radiation’ tends to be emitted from, and fly away from, the ‘charged stuff’, then one expects, in line with general expectations (see [6] and references therein) that the time-evolution will be approximable by the action on $\rho_m(0)$ of a *semi-group* of completely positive maps whose exact solutions obey a master equation of GKS/L form. (This semigroup will extend to act on arbitrary – i.e. not-necessarily-pure $\rho_m(0)$) We remark that, although this is not always mentioned or emphasized, one of course equally expects the time evolution for negative times to be similarly governed by a master equation of ‘anti-GKS/L’ form (i.e. a generator of a completely positive semigroup in the variable $t' = -t$) and this will obviously be the case if the Hamiltonian H and initial state $\Psi(0)$ mentioned above are both time-symmetric in obvious suitable senses.

Benatti and Narnhofer [36] have given necessary and sufficient conditions for the von Neumann entropy of a density operator evolving in time (for positive times) according to a master equation of GKS/L form to increase monotonically. One can easily check that their conditions hold for BLP and we mention in passing that it is shown in [36] that they hold for the time-evolution of GRW [13]. (The same result with S replaced by S_1 – defined in Section 6 – is demonstrated in Section 6 for BLP and in [13] itself for GRW.) When a total unitary time evolution as in the previous paragraph leads to an approximate GKS/L-form master equation governing the behaviour of ρ_m at positive times which satisfies Benatti and Narnhofer’s conditions, then one clearly expects entropy to increase for positive times. Although this is again not always mentioned or emphasized, one also expects (and again of course this must happen if the total dynamics and time-zero state has a time-reversal symmetry) the entropy of ρ_m will increase for increasingly negative times. We shall call such an overall time-behaviour for the von Neumann entropy of ρ_m as t ranges over the whole real line a ‘two-sided entropy increase’ result.

(The above two paragraphs are particularly relevant to the discussion in Section 6.)

- (x) To see that the solutions to (9) and (10) specified by the formulae (8) and (6) consist of density operators at all times, we note that it is clear from the fact that $D(0) = 1$ that ρ will have unit trace. It is also clear from the way that (3) and

(8) etc are derived that $\rho(t)$ will be a positive operator at all t , and this can be shown to hold also when D in (8) is replaced by either its Gaussian or logarithmic asymptotic forms.

The failure of the master equations (9) and (10) to have GKS/L form can be traced back to the fact that the modes which are traced over in obtaining ρ_{physical} (see Note <vi>) are, in the Newtonian case, non-radiative and, in contrast to what was envisaged for ‘radiation’ in our discussion of the approximate validity (for positive times) of master equations of GKS/L form in the situation envisaged in Note <ix>, instead of ‘flying away’ from our matter, they are ‘dragged around with’ (or one might say ‘slaves to’) the matter. (Cf. the last paragraph of our discussion of the Penrose experiment in Subsection 4.1 and Note <xxv> and also (especially the last paragraph in) Section 6.

Below, we shall call time-evolutions on spaces of density operators of form described in the paragraph containing equation (8) ‘multiplicative decoherence models’ (MDM) and master equations such as (9) and (10) ‘MDM master equations’.

It is interesting to compare and contrast MDM master equations with master equations of GKS/L form (see Note <ix> and Section 6). MDM master equations have solutions for all times, t , ranging over the real line while GKS/L master equations only have solutions for positive t . On the other hand, for positive t , GKS/L master equations have solutions for arbitrary initial density operators while MDM master equations only have solutions for very special initial data (but including of course initial data of form (3) for equation (9) and (6) for equation (10) etc.) What goes wrong for GKS/L equations for negative times, and for MDM master equations for wrong initial data is that one can find solutions in the class of unit-trace trace-class operators, but they will fail to stay in the class of density operators because they will fail to satisfy positivity. Related to this, one can prove (BSK unpublished), for our Gaussian model master equation (10), the theorem:

The time evolve after an arbitrarily small positive time δt of a pure density operator $\rho = |\psi\rangle\langle\psi|$, where ψ is even, fails to be positive except when ψ takes the Gaussian form $\psi = e^{-cx^2}$ with $\text{Re}c > 0$ and $\text{Im}c < 0$. (In the special case with $c = 4\kappa$, $\rho = |\psi\rangle\langle\psi|$ evolves into itself for all later times.)

If, however, we compare and contrast MDM models, not with time-evolutions generated by master equations of GKS/L form, but rather with time-evolutions such as the sort described in <ix> which, as we discussed in that note, are expected to be approximately of GKS/L form for positive times *and of anti-GKS/L form for negative times* then, as far e.g. as the behaviour of entropy is concerned, the two sorts of models can have qualitatively similar behaviour. Namely, in both cases (depending on the details of the dynamics) one can have a two-sided entropy increase result in the sense explained in <ix>. We show such a two-sided entropy-increase result for our Gaussian (MDM) model with a free Hamiltonian (and to first order in κ) in Section 6.

- (xi) As we shall discuss further in Section 4, we expect that our conclusions will also apply if the probe particles are relativistic (e.g. photons).
- (xii) The purpose of this Note is to sketch a natural extension of the hypothesis of [1] (see subsection 1.1 of the Introduction and also Notes <i> and <v>) to deal with open systems. In this extension, the specification of an open subsystem of some given closed system is taken to correspond to a particular way of expressing the Hilbert space $\mathcal{H}_{\text{matter}}$ as a tensor product

$$\mathcal{H}_{\text{matter}} = \mathcal{H}_{\text{matter, system}} \otimes \mathcal{H}_{\text{matter, environment}}$$

so that, in view of (1), the Hilbert space, $\mathcal{H}_{\text{total}}$ for the total system will arise as a triple tensor product

$$\mathcal{H}_{\text{total}} = \mathcal{H}_{\text{matter, system}} \otimes \mathcal{H}_{\text{matter, environment}} \otimes \mathcal{H}_{\text{gravity}}.$$

For a given ρ_{total} (which, see (2), will, according to our hypothesis, take the form $|\Psi\rangle\langle\Psi|$) of the relevant total closed system, the density operator, $\rho_{\text{matter, system}}$ describing the partial state of the matter belonging to the open subsystem is then declared to be the partial trace of ρ_{total} over $\mathcal{H}_{\text{matter, environment}} \otimes \mathcal{H}_{\text{gravity}}$ and the physical entropy of the open subsystem is declared to be the von Neumann entropy of $\rho_{\text{matter, system}}$. All this may be represented by the schematic rectangle-picture in Figure 4a where one may regard the vertical dividing line as separating the matter degrees of freedom which belong to the ‘open subsystem’ (marked ‘matsys’ on the figure) from those which belong to its ‘environment’ (marked ‘matenv’ on the figure) while the horizontal line separates all matter from gravity (indicated by the region marked ‘grav’). In the sense that it is always part of what is traced over, one might say that gravity is regarded as a *permanent part of the environment* so that the horizontal line is fixed, while the vertical line is slideable. Sliding it to the right corresponds to gradual enlargement of our open subsystem of interest until, in the limit as it coincides with the right hand boundary of the lower rectangle, the open system will approach the total closed system and its entropy will approach the entropy of the total closed system which, of course, will still be non-zero because all of $\mathcal{H}_{\text{gravity}}$ continues still to be traced over in the definition of that. Indeed, if our total closed system contains a (say stellar or galactic-centre sized) black hole, then this limiting value of the entropy will be very large. As we indicate in the schematic graph Figure 5a, it seems reasonable to expect that the entropy of ever larger (say nested) subsystems will increase monotonically with the size of the subsystems.

All this is in contrast to, and we feel, more satisfactory than, the situation if one were to attempt a definition for the entropy of an open subsystem of a given closed system on traditional ideas (i.e. on the assumption that the gravitational field due to the matter of the subsystem is regarded as belonging to the subsystem) by defining it to be the von Neumann entropy of the partial trace of the total density operator over the Hilbert space for the total environment of the subsystem, (and assuming the total density operator to be pure). In other words, by defining it to be the entanglement entropy of the subsystem with its total environment (and

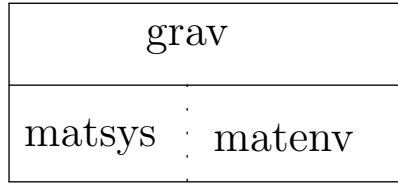


Fig. 4a

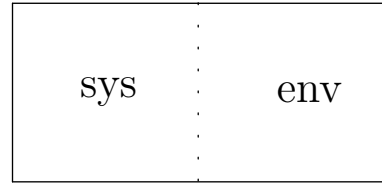


Fig. 4b

Figure 4. Schematic diagrams contrasting our approach to open systems (Fig. 4a) with that on the traditional ‘environment-induced decoherence’ paradigm (Fig. 4b)

assuming the total density operator to be pure). This might be represented by the schematic rectangle figure 4b. The entropy, thus defined, for a given open subsystem would then necessarily (cf. Note <iv>) equal the entropy of its environment and therefore, as one considers ever larger nested subsystems which eventually approach the total closed system, as indicated in figure 5b by sliding the vertical dividing line to the right, the adage ‘what goes up must come down’ would necessarily apply and the entropy would approach the value zero! So a typical graph of entropy against size of region would, in contrast to Figure 5a look like that sketched in Figure 5b.

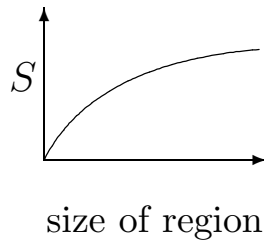


Fig. 5a

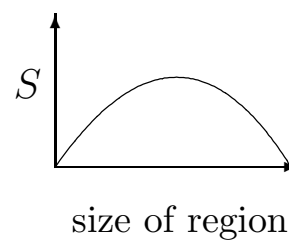


Fig. 5b

Figure 5. Schematic behaviour of entropy against ‘size of open system’ for our approach to open systems (Fig. 5a) contrasted with the corresponding behaviour on traditional ideas (Fig. 5b)

Finally, we remark that, in defining the density operator of any open system, the result of tracing over $\mathcal{H}_{\text{matter, environment}} \otimes \mathcal{H}_{\text{gravity}}$ is of course equivalent to first tracing over $\mathcal{H}_{\text{gravity}}$ and then tracing over $\mathcal{H}_{\text{matter, environment}}$. So our proposal is equivalent to a modification of the above-described approach to defining entropy on traditional lines – i.e. as pictured in Figure 4b – in which one identifies ‘system’

as ‘matter system’ and identifies ‘environment’ as ‘matter environment’ but, instead of postulating a total state which is pure, one postulates a total state on the full matter system which is equal to the (mixed) trace of our matter-gravity ρ_{total} over $\mathcal{H}_{\text{gravity}}$.

- (xiii) In this note, we shall add a number of further remarks about our resolution to the Schrödinger Cat puzzle discussed in Sections 1 and 5:

We begin by remarking that, in the context of Schrödinger Cat-like situations, something similar to our mechanism whereby a pure state such as (11) gets changed to a mixed density operator like (68) or (approximately – see below) like (12) occurs, of course, in the general context of ‘environmental induced decoherence’ (see e.g. [25] and [26] and references therein). However there are a number of important differences. First, the ‘environmental induced decoherence’ mechanism is only capable of bringing about the replacement of a pure by a mixed state in the case of an *open* system, whereas in the theory of [1] and [2], by having a preferred environment (i.e. gravity) and declaring that this must always be traced over in principle, such a replacement of pure by mixed states occurs for closed systems. Secondly, something similar to our proposed general interpretation (see around the passage labelled (B) in Section 1 and the beginning of Section 5) of our physical density operators in terms of events which happen is usually implicitly or explicitly assumed when discussing the interpretation of density operators of form (12) – but it usually seems to be assumed that the events which can happen should be identified with *one-dimensional* diagonalising subspaces of the relevant density operator (and that the probability with which they happen is the associated eigenvalue). In the case of (12) this will not change the interpretation as long as $|c_1|^2 \neq \frac{1}{2}$. But in that special case, on the latter usual interpretation, an ambiguity occurs since the eigenvalue $\frac{1}{2}$ is degenerate. The existence of this ambiguity in this usual interpretation is sometimes taken as a reason to question whether a formalism based on density operators can really solve the conceptual problems of quantum mechanics. However our interpretation (B) sidesteps this criticism by entailing that, in this same special case, there is only *one* possible event which can happen (with probability 1), namely that corresponding to the subspace spanned by ψ_1 and ψ_2 . This is free from ambiguity although it is admittedly somewhat strange. However, the strangeness is perhaps mitigated in that, in a typical preparation of a would-be Schrödinger Cat-like state, such as (11), the quantities c_1 and c_2 would be expected to vary in time, and the equalities $|c_1|^2 = \frac{1}{2} = |c_2|^2$ to only hold for a single moment of time, and hence our interpretation would entail that there are two possible events, each evolving in time, then fusing into one possible event for a fleeting moment before splitting into two (evolving) possibilities again. The ball (or the cat!) is still predicted to be one thing or another except at that moment and at no time is the statement about the possible events which can happen (and the probabilities with which they will happen) ambiguous.

Finally we remark that $\langle g_1|g_2 \rangle$ in (68) can be small but never actually zero and therefore the density operator (12) can only be an approximation to (68) in Note <vi>. In fact, in general the subspaces belonging to our two events will clearly be one-dimensional subspaces spanned by (normalizable) vectors of the form

$$\phi_1 = k\psi_1 + s^*\psi_2 \quad \text{and} \quad \phi_2 = -s\psi_1 + k^*\psi_2 \quad (70)$$

with $|k|^2 + |s|^2 = 1$ and ρ will take the form

$$A|\phi_1\rangle\langle\phi_1| + B|\phi_2\rangle\langle\phi_2| \quad (71)$$

where A and B are positive numbers with sum 1 and the values of A , B , k and s will be determined (by equating (71) with (68)) in terms of $|c_1|^2$, $|c_2|^2$ and $\langle g_1|g_2 \rangle$. When, the ball is of mass much smaller than the Planck mass and/or the two ball states are close together, $\langle g_1|g_2 \rangle$ will be close to 1, A will be close to 1, B close to zero, k close to c_1 and s close to c_2 so there will be one ‘event’ with a probability close to 1 which is close to the subspace spanned by ψ of (11) while, at the other extreme, returning to the case of almost complete decoherence, as our ball gets larger and/or as the two ball-states get further apart, k will be close to 1 and s close to zero, and A will be close to $|c_1|^2$ and B will be close to $|c_2|^2$ and k will get closer to 1, and s closer to zero. We feel that our statement that this “is just what one would hope for from a resolution to the Schrödinger Cat puzzle” remains equally arguable with this exact description of the event-subspaces.

- (xiv) As far as we are aware, an interpretation in terms of ‘events which happen’ along the lines of that proposed here in (B) has not been considered previously in the context of models for modified quantum mechanics which are couched in terms of time-evolving density operators. However, something similar to (but not quite identical to) our proposal seems often to be implicitly assumed when considering the physical interpretation of density operators which arise in contexts such as proposed resolutions to the Schrödinger Cat puzzle – see Note <xiii>.
- (xv) Here, we reserve the term ‘collapse models’ for theories such as the ‘GRW’ model [13] when these are formulated in terms of time-evolving density operators. For a given such collapse model, (and provided, as is e.g. the case for the GRW model, the time-evolving density operator arises from a semigroup of completely positive maps – see Note <ix> – acting on some initial density operator) it is now understood that there are typically several different ways of representing the dynamics in terms of a time-evolving stochastic wave function – each constituting what is (since the term was coined around 1993 by Carmichael [15]) now called a different ‘unravelling’ (see e.g. [14]) of the model. An early example is the ‘usual’ unravelling of the GRW model which actually predates the use of the term ‘unravelling’. This was formulated in a simple explicit way by Bell in [16]. For a given collapse model, a choice of unravelling often suggests another way to that proposed here of interpreting the collapse model in terms of ‘events’ which ‘happen’. For the ‘Bell’ unravelling of the GRW model, this is explained in [16]. The proposal made

here for how a notion of ‘events which happen’ may be obtained from a given time-evolving density operator is offered as an alternative to such interpretations (see Note <xiv>). It involves a more abstract notion of ‘events’ to the ‘collapse centres’ of Bell/GRW. It may suffer from a number of problems, but it is free from the problem of the ambiguity of unravellings. In any case, the notion of unravellings is only available for time-evolutions which *exactly* arise from semigroups of completely positive maps – i.e., cf. <ix>, which obey master equations of exact GKS/L form. Hence, unravellings are anyway unavailable to us since the master equations, (42) and (43), of relevance to us are not of this form. (At best, (43) is *approximately* of this form (for positive times). For similar reasons, unravellings are unavailable for more traditional situations where one has a time-evolving density operator arising from some sort of environment-induced decoherence ([25], [26]) since such time-evolutions also do not constitute semigroups of completely positive maps (these again only arising, if at all, after some sort of mathematical limiting procedure). On the other hand, the events interpretation we propose here would be available.

- (xvi) It is of course a tricky philosophical issue just what one means by ‘probability’ here. (Some common views about what probabilities are make reference to observers!) We will not attempt to say anything original about this issue here, but we note that more or less the same issue is present with notions of events such as that proposed by Bell for the GRW model (see Note <xv>). A recent article which addresses the issue in this latter context is [17].
- (xvii) The statements about the way the pencil gets reflected off the mirror can easily be verified with a simple conservation of momentum argument making the usual small-angle approximations. In explanation of the various assumptions: (a) the diameter of the pencil is assumed to be much bigger than δ in order to ensure that the angle of widening of the pencil due to spreading of the wave packet will be much less than the angles between the various beam components after the reflection. (There will also be an initial broadening of all the reflected pencils by an amount $\sqrt{2}\delta$ because of the uncertainty in the location of the mirror.); (b) the diameter of the mirror has obviously to be large enough for the pencil to be entirely reflected by it, no matter where along the interval $(0, 2\delta)$ of the x -axis the centre of mass of the bead is located; (c) the mass of the probe particles is taken to be much smaller than M so that reflection by the mirror will consist simply in reversal of the perpendicular component of momentum in the centre of mass frame; (d) the assumption $P \gg p$ was made simply so that we can use the usual small angle approximations to obtain the uniform spacing in the angles of the reflected beam components.
- (xviii) A similar declaration to that considered in the discussion motivating our *corrected pragmatic interpretation* in Section 4 has been made before in the context of other approaches to the measurement problem in quantum mechanics. Cf. e.g. the discussion around the following passage in the article [11] by John Bell concerning what might constitute what Bell calls a ‘beable’: “Not all ‘observables’ can be given

beable status What is essential is to be able to define the positions of things, including the positions of instrument pointers or (the modern equivalent) of ink on computer output.”

- (xix) One could e.g. presumably undo the measurement of the bead momentum inserting a suitable lens-like device between the probe particle and the screen so that, however its wave function reflected off the mirror, only one spot forms on the screen.
- (xx) The version described here is similar to the ‘FELIX’ experiment described in [21] and to the ‘suggested space-based experiment’ described in Appendix 2 of [22]. In the realistic versions of the experiment described in these references, Penrose envisages the photon to be an X-ray photon, and the movable mirror to be a Mössbauer crystal. Penrose and collaborators have also proposed a ground-based experiment (see again Appendix 2 in [22] and [9]) in which the photon and movable mirror are optical and the photon is stored in suitable optical cavities. The details of the ground-based experiment (see [9]) are slightly different but it is easy to see that our discussion here and in the following subsection could easily be adapted to this latter experiment and the conclusions we draw here would survive in essentially unaltered form. We remark that what is difficult about these experiments (and is discussed in the references just given) is of course to eliminate all causes of decoherence other than the possible cause that the experiments are designed to test.
- (xxi) We have made some extra implicit assumptions and simplifications here. In particular, the state of the photon/movable-mirror-cum-oscillator system at time t_m might well be more accurately modelled by an entangled state, rather than the single tensor product we have assumed here. Without extending our analysis to include such a possibility, we note that we are still assured by our position measurement theorem that, irrespective of how we model it, as long as the Penrose experiment ends in a type of position measurement (and assuming that photons may be treated for this purpose as non-relativistic particles) the prediction of the theory of [2] will still be identical with the prediction of standard quantum mechanics.
- (xxii) We should really have allowed for a relative phase in the two equations (28), (29). However, for simplicity we assume the phase happens to be 1 and that, with this phase, one predicts zero detection rate at the detector. If either of these were not the case, one could e.g. imagine modifying the experiment by inserting a suitable phase-shifting device in one arm of the interferometer in order to continue to predict zero detection rate at the detector.
- (xxiii) In his 1996 article, [20], Penrose explains that what he actually envisages is a radical new theory of ‘quantum state vector reduction’ and he emphasises that the technical form that such a theory will take is not yet known. Rather, some qualitative and semi-quantitative arguments are given and the direction in which such a theory may lie, indicated. In our discussion of the Penrose experiment in Subsection 4.1 we perhaps should have written “Now what is envisaged by Penrose, to the extent that it may be represented within current quantum mechanical formalism involving density operators etc. . . .”.
- (xxiv) Interestingly, we easily see that the partial state of the photon at time t_m is predicted to be the same as that of (30) by all three theories:

$$\rho_{\text{matter}}^{\text{Kay}}(t_m) = \rho_{\text{matter}}^{\text{Penrose}}(t_m) = \rho_{\text{matter}}^{\text{standard}}(t_m)$$

(However, in the Penrose experiment, no direct measurement is made of this state.)

- (xxv) We remark that it is of course the equality of $g(t_f)$ with $g(t_0)$ (or more generally the fact that, at any time, t , $g(t)$ depends only on the matter-configuration at time t) in the non-relativistic case which is responsible (cf. Note <viii>) for the form of (the relevant generalization to the photon/movable-mirror-cum-oscillator system of) equation (8) which (together with the relevant version of equation (27)) is responsible for the validity of our Position Measurement Theorem (when applied to the Penrose experiment).

Further insight into the reasons for the recoherence may be had by considering the simple model closed system consisting just of a movable-mirror-cum-oscillator as in the Penrose experiment, but uncoupled from any photons etc. The centre-of-mass motion of such an oscillator is of course modelled in standard quantum mechanics as a standard quantum-mechanical harmonic oscillator. Any (Schrödinger picture) wave-function solving this will, of course, be periodic in time and taking the would-be density operator, $\rho_0(t)$, in (a one-dimensional version of) (8) to be $|\psi(t)\rangle\langle\psi(t)|$ for such a periodic would-be wave function, it is clear, from (8), that ρ will also be periodic and hence alternately decohere and recohere. This is in contrast with the typical prediction of a ‘collapse model’ – see Section 6 and Notes <ix> and <x> for further discussion.

- (xxvi) By Einstein’s quadrupole formula, we would expect the energy radiated per unit time to be given by a number of order 1 times $m^2\ell^4\omega^6$ (if we restore c and G , this should be multiplied by G/c^5) where m is a relevant mass (assuming the mass of the ‘oscillator’ part of the movable-mirror-cum-oscillator [the ‘cantilever’ in [9]] can be neglected, then this would be the mass of the movable mirror and its mounting) and ℓ its amplitude, and ω its frequency of oscillation. Thus the expected number, N_{graviton} , of gravitons emitted per cycle would be a number of order 1 times $m^2\ell^4\omega^4$ (times $G/(\hbar c^5)$) (times 2π). Putting in the values (similar to or larger than those in the experiment suggested in [9]) $m = 10^{-8}$ gram, $\ell = 10^{-11}$ cm, $\omega = 3 \times 10^3 \text{ sec}^{-1}$, this gives an N_{graviton} of around 10^{-80} which is utterly negligible!

We remark that there will actually inevitably be another source of decoherence due to a form of radiation by the oscillating mirror, namely the radiation by the movable mirror of photons at around the oscillator frequency due to the effect of the oscillating boundary condition imposed on the quantized electromagnetic field by the oscillating mirror. Ford and Vilenkin [23] have calculated that, for such an oscillating mirror, the energy per unit time radiated by this mechanism for a scalar analogue of the photon is given by (note that a missing factor of 2π in [23] has been inserted here) $E = \frac{1}{720\pi^2}\ell^2\omega^6 A$ (times \hbar/c^4) where A is the surface area of the mirror, and we shall assume that the same formula, multiplied by 2 because of the photon’s two polarization states, holds true for true photons. (It is not actually known whether or not this assumption is correct, but it will surely give the correct order of magnitude.) The expected number, N_{photon} , of photons emitted per cycle will then be $\frac{1}{180\pi}\ell^2\omega^4 A$ (times $1/c^4$). With the same values assumed above for ℓ and ω and with a value for A (again taken from the suggested experiment [9]) of 10^{-6} cm^2 , this gives an N_{photon} of around 10^{-39} which is vastly bigger than N_{graviton} but of course, for the purposes of the Penrose experiment (in the version proposed in [9]), reassuringly, still vastly smaller than anything which would need to be taken into account in a practical analysis of the experiment.

- (xxvii) The discussion of the relationship between the notion of parity in the naive pragmatic interpretation and the notion of parity in our events interpretation and, in particular, the coincidence explained after equation (36) may clearly be generalized as we now explain: It seems natural to introduce the following definitions and terminology: For any ρ to

which one intends to apply our events interpretation, we say that a self adjoint operator, B , is a *beable* for ρ if it commutes with ρ and if all spectral subspaces of ρ belonging to non-zero eigenvalues of ρ are also spectral subspaces of B . In other words, a beable for ρ is a self-adjoint operator which commutes with ρ and, on the orthogonal complement of the kernel of ρ , is a function of ρ . We then say that an eigenvalue of B on a given event (i.e. spectral subspace) of ρ is called the *value* of the beable B when that event occurs.

With this definition, it is obvious that, for a given ρ , a similar coincidence to that remarked upon after equation (36) will occur quite generally between the expected value of a beable, B , in the sense of *the sum over all possible events of the probability of each event which can occur multiplied by the value of the beable B when that event occurs* and the expectation value of B in the conventional sense of $\text{tr}(\rho B)$.

We have taken the word ‘beable’ from the writings of John Bell – see e.g. [11] where Bell writes:

“*The beables of the theory are those elements which might correspond to elements of reality, to things which exist. Their existence does not depend on ‘observation’. Indeed observation and observers must be made out of beables.*”

However, we should caution that the candidate for a beable being proposed here differs from various candidates that Bell himself proposed (see e.g. Note <xviii>) and, in particular, we would remark that, in contrast to other proposals for beables, in our notion, what is a beable at a given time depends on the state of the system at that time.

A couple of interesting features of our notion of beable are worth pointing out. First if, in the case of an n -dimensional Hilbert space, ρ is $1/n$ times the identity operator then the only beables will be multiples of the identity operator. This applies, for example to the Schrödinger Cat state (12) when $|c_1|^2 = \frac{1}{2} = |c_2|^2$. (See also the discussion of this special situation in Footnote <xiii>.) At the other extreme, if ρ arises as the projector $|\psi\rangle\langle\psi|$ onto some vector ψ , then any self-adjoint operator which commutes with ρ is a beable.

References

- [1] Kay BS 1998 Entropy defined, entropy increase and decoherence understood, and some black-hole puzzles solved *Preprint* hep-th/9802172
- [2] Kay BS 1998 Decoherence of macroscopic closed systems within Newtonian quantum gravity *Class. Quantum Grav.* **17** L89-L98 (*Preprint* hep-th/9810077)
- [3] Abyaneh V and Kay BS 2005 The robustness of a many-body decoherence formula of Kay under changes in graininess and shape of the bodies *Preprint* gr-qc/0506039
- [4] Abyaneh V 2006 *Gravitationally Induced Decoherence* University of York PhD thesis
- [5] Wald RM 1998 Black holes and thermodynamics, In *Black Holes and Relativistic Stars* Ed. Robert M. Wald (Chicago: Chicago University Press)
- [6] Gorini V, Kossakowski A and Sudarshan ECG 1976 Completely Positive Dynamical Semigroups of N -Level Systems *J. Math. Phys.* **17** 821-825
- [7] Lindblad G 1976 On the generators of quantum dynamical semigroups *Commun. Math. Phys.* **48** 119-130
- [8] Folman R, Schmiedmayer J and Ritsch H 2001 On the observation of decoherence with a movable mirror *Z. Naturforsch. A* **56** 140-144 (*Preprint* quant-ph/9906064)
- [9] Marshall W, Simon C, Penrose R and Bouwmeester D 2003 Towards quantum superpositions of a mirror *Phys. Rev. Lett.* **91** 130401 (*Preprint* quant-ph/0210001)
- [10] von Neumann J 1983 *Mathematical Foundations of Quantum Mechanics* (Princeton: Princeton University Press)

- [11] Bell JS 2004 Beables for quantum field theory. In: Bell JS *Speakable and Unspeakable in Quantum Mechanics* Cambridge University Press (originally published in 1984 as CERN report CERN-TH 4035/84)
- [12] Bell JS 1990 Against ‘measurement’. In *Sixty-Two Years of Uncertainty. Historical, Philosophical and Physical Inquiries into the Foundations of Quantum Mechanics*, Proceedings of a NATO Advanced Study Institute, 5-14 August 1989 Erice, Miller, AI, editor, NATO ASI Series B vol. 226, Plenum Press, New York (Also published as *Physics World*, vol 3, August 1990, pages 33-40 and also reprinted in the second (2004) edition of Bell JS *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press))
- [13] Ghirardi GC, Rimini A and Weber T 1986 Unified dynamics for microscopic and macroscopic systems *Phys. Rev. D* **34** 470-491
- [14] Gisin N, Brun TA and Rigo M 1996 From quantum to classical: The quantum state diffusion model, In *New Developments on Fundamental Problems in Quantum Physics: Proceedings 1997* Ferrero M and Van der Merwe A, editors (Dordrecht, Netherlands: Kluwer) (Fundamental Theories of Physics, Vol. 81) (*Preprint* quant-ph/9611002)
- [15] Carmichael HJ 1993 *An Open Systems Approach to Quantum Optics* Lecture Notes in Physics (Berlin: Springer-Verlag)
- [16] Bell JS 1987 Are there quantum jumps? In: *Schrödinger: Centenary Celebration of a Polymath* Kilmister CW, editor (Cambridge: Cambridge University Press) (pages 41-52). (Also reprinted in Bell JS 2004, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press))
- [17] Frigg R and Hoefer C 2007 Probability in GRW theory *Studies in the History and Philosophy of Modern Physics* **38** 371-389 (*Preprint* <http://philsci-archive.pitt.edu/archive/00003160/>)
- [18] Adler RJ and Santiago DI 1999 On gravity and the uncertainty principle *Mod. Phys. Lett. A* **14** 1371 (*Preprint* gr-qc/9904026)
- [19] Arndt M, Nairz O and Zeilinger A 2002 Interferometry with macromolecules: quantum paradigms tested in the macroscopic world. In *Quantum [Un]speakables* Bertlmann RA, Zeilinger A, editors (Berlin: Springer Verlag)
- [20] Penrose R 1996 On gravity’s role in quantum state reduction *Gen. Rel. Grav.* **28** 581-600
- [21] Penrose R 2000 Wavefunction collapse as a real gravitational effect, in *Mathematical Physics 2000*, Fokas A et al, editors (London, Imperial College)
- [22] Penrose R et al 2000 *The Large, the Small and the Human Mind* (‘Canto’ paperback edition) (Cambridge: Cambridge University Press)
- [23] Ford LH and Vilenkin A 1982 Quantum radiation by moving mirrors *Phys. Rev. D* **25** 2569-2575
- [24] Wehrl A 1978 General properties of entropy *Rev. Mod. Phys.* **50** 221-260
- [25] Zurek WH 1991 Decoherence and the transition from the quantum to the classical *Physics Today* **44** 36-44
- [26] Joos E, Zeh HD, Kiefer C, Giulini D, Kupsch J and Stamatescu I-O 2003 *Decoherence and the Appearance of a Classical World in Quantum Theory* (Berlin: Springer Verlag)
- [27] Barchielli A, Lanz L and Prosperi GM 1982 A model for the macroscopic description and continual observations in quantum mechanics *Nuovo Cimento* **72B** 79
- [28] Daneri A, Loinger A and Prosperi GM 1962 Quantum theory of measurement and ergodicity conditions *Nuclear Physics* **33** 297-319. (Also reprinted in Wheeler JA and Zurek WH 1983 *Quantum Theory and Measurement* (Princeton: Princeton University Press))
- [29] Sewell G 2005 On the mathematical structure of quantum measurement theory *Rep. math. Phys.* **56** 271 (*Preprint* quant-ph/0505032)
- [30] Reichenbach H 1971 *The Direction of Time* (Berkeley: University of California Press)
- [31] Gibbons GW and Hawking SW 1977 Action integrals and partition functions in quantum gravity *Phys. Rev. D* **15** 2738-2756
- [32] Hawking SW 1979 The path integral approach to quantum gravity, in *General Relativity: An Einstein Centenary Survey* (Cambridge: Cambridge University Press)

- [33] Bekenstein J 1973 Black holes and entropy *Phys. Rev. D* **7** 2333-2346
- [34] Hawking SW 1975 Particle creation by black holes *Commun. Math. Phys.* **43** 199-220
- [35] Page D 2005 Hawking radiation and black hole thermodynamics *New J. Phys.* **7** 203 (*Preprint* hep-th/0409024)
- [36] Benatti F and Narnhofer H 1988 Entropy behaviour under completely positive maps *Letters in Mathematical Physics* **15** 325-334